Final Report – Evaluation of the Efficacy of Futures by Design

William Davies, Magdalena Partac, Emanuele Giovannetti, Wei Kang, Marcus Bellett-Travers
ANGLIA RUSKIN UNIVERSITY, December 2022

Contents

Introduction

The purpose of this report is to evaluate the efficacy of the Futures By Design (FBD) project across Seven North Sea EU regions, providing both a qualitative and quantitative analysis of the project outcomes. The FBD project worked with 360 SMEs across five European countries: The Netherlands, Belgium, the United Kingdom (UK), Germany and Sweden, with individual projects developed for each region.

The qualitative analysis focuses on considering key case studies from each of the regions, before drawing upon these case studies to identify key features and themes for further discussion and investigation. Whereby the quantitative section considers whether we can determine the internal (SME characteristics) factors that have impacted the completion of data projects. Thus, this report consists of the following five sections: Introduction, Case Studies, Discussion from themes identified in the case studies, Quantitative analysis, Conclusions and Recommendations.

Before considering the assembled evidence the authors briefly introduce the key theories and theoretical concepts of *Potential* and *Realised Absorptive Capacity* which will be discussed and utilised both in the qualitative and quantitative analysis.

Potential and Realised Absorptive capacity

Evolutionary economists have shown that technological development and innovation capability are significant drivers of the evolution of the industrial structure (Dosi and Nelson, 2018; Krüger, 2008; Nelson, 2020). Building on the Schumpeterian theory (Schumpeter 1911) that entrepreneurship and technical change are at the core of the economic growth process, these economists contend that knowledge acquisition and dissemination are the primary factors influencing how enterprises' internal capabilities develop and how industrial structures as a whole change over time.

These concepts are adapted from Jansen, Van den Bosch, and Volberda (2005), which, in turn, were based on Zahra and George (2002) and Szulansky (1996). *Potential Absorptive Capacity* (PAC) captures the ability for an enterprise to identify and assimilate external knowledge relating to innovation and technological advancement. Conversely, *Realised Absorptive Capacity* (RAC) identifies a firm's ability to actualize the assimilated external knowledge into the firm. For example, a firm demonstrates *Potential Absorptive Capacity* by identifying that a Customer Relationship Management system (CRMs) could be useful in managing its clients, while its realised absorptive capacity is demonstrated if a firm implements such a system.

Data Maturity

One of the key concepts utilized throughout this project was that Data maturity was a process by which we identified the level of usage of data within the SME, this was achieved through utilizing an in-depth questionnaire and weighing the answers to calculate data maturity on a 1 to 5 scale. This was considered a vital onboarding step for the SMEs as it enabled the identification of what sort of data projects were plausible. The diagram below demonstrated the different levels of data maturity within the project:

Data Maturity Scale

1 Beginner	No experience at all with data, starting from scratch
2 Gatherer	Starting to collect data, but has not starting analysing
3 Describer	Has data and is analysing it but not using it to predict outcomes
4 Predictor	Is using Data and analysing data but has some areas to expand
5 Expert	Is an expert in all aspects of data analysis and collection

Figure 1 Data maturity scale

Thus, for every single SME, the data maturity level at which the company started was identified. This was a key step in enabling a reasonable data project to be selected as complex data projects could only be achieved with high data maturity, therefore the designed project had to reflect the existing level of data maturity.

Section 1 Case Studies by region

The following section presents a selection of case studies for each of the regions involved in the study, before considering key themes identified in each region. The section is subdivided by country.

Sweden

Project Question	Goal	Data	Use of Data	Business Value
Can we use data to improve the service level instead of the feel-good conversations that we are currently having with our B2B customers?	At some point, they hope to ensure that their clients remain devoted to them and deepen their professional relationships.	There was no data yet, so a questionnaire is created with questions about the service level and how satisfied they are with the service itself.	Analyze the questionnaire results to understand what customers are happy with and find development areas and suggestions for new services.	Eventually, this led to extended customer relationships and additional sales.
We are aware that some of the items in our warehouse are rented either infrequently or never. Can we manage our warehouse's inventory more effectively?	Learn more about how often people rent various products. To reduce the number of unprofitable products, generate extra inventory and capital, and reinvest the proceeds back into our company.	Information from the warehouse system that reveals how frequently a product is rented.	Find out what is frequently or infrequently rented by analysing the data.	More money is needed to enable business/new product reinvestment.

Is it possible to allow everyone to identify a place to live, where their needs and personalities are best met?	Make a database and search engine for Swedish locations. For people to find "their best spot," information about locations can be retrieved using the same phrases and search terms.	There were no data yet. Therefore, qualitative interviews were developed in which respondents were asked about their personalities and preferences.	Following an analysis of the qualitative interview responses, various categories for various types of persons are developed.	Utilising qualitative data to increase the profit levels of the business.
Can the company structure its online platform to effectively gather more structured data from users and thus increase the sustainability of events?	Better and more engaging events should be created for the audience. Exploring the needs of the clients in the process.	Since there was no data at the outset, the first step is to gather data.	Data reports are created to provide deeper insight into the needs of the customer.	In the end, they gave the client superior service. Additionally, more information about the industry's interest is generated.

Key themes from the Swedish case studies:

- Firms with no data at all:

One of the key notes of these case studies is that over half of the examined case studies started with no data. If through the lens of absorptive capacity, we could consider that without data it may be more difficult for a firm to identify areas upon which it could improve its technical ability.

- Improving customer relationships, a key goal for multiple studies:

Multiple the Swedish case studies note that companies wish to use data in order to improve their customer relationship.

- Questionnaires and surveys:

This was utilised multiple times as a first key step in these case studies to start the data maturity journey.

- Existing unused data:

One of the key themes here is that many firms had data but simply had not analysed the data.

United Kingdom

Project Question	Goal	Data	Use of Data	Business Value
The Yoga studio began offering its members internet classes during COVID. Is it still beneficial to attend online classes now that COVID is less prevalent, and the requirements have been softer?	Evaluating the organization's membership data to evaluate the viability of ongoing online courses.	A large amount of data was collected from an online booking system, including numbers of attendance and whether they were online or offline.	The data was used to decide whether they should continue to offer online courses.	The decision was made to keep the online courses as over 50 percent of their current revenue was from online courses.
What information about each farm does an Agritech firm need	Expanding and taking into account the information	Exceptionally large datasets, yet	Data informed what questions	More knowledge about the data they would need to gather

to collect in order to make business decisions.	that would be useful when taking into account the business problems for each customer.	just a relatively small amount of business information gathered from clients.	should be asked of the farms in order to create a more succinct set of datasets.	to make important business decisions in the future. They were able to grasp the data's intended use thanks to the framing mechanism.
What new business prospects in the vertical farming sector can data identify?	Based on their preferences and technological capabilities, assist them in identifying new directions for the company.	They make use of data on their pre-existing connections.	Their pre-existing data was used to identify potential new business opportunities focused on expansion of their base products	This process enabled the business to identify niche markets and thus begin the process of expanding its business.
How can employee practice information be used to help build an online training database?	We concentrated on how their continuing data analysis may be refocused on supporting their online training product given their extensive	The business has extensive data about its staff members and training. And they possess the advanced technical	They used the data to help them inform the construction of a new training hub.	identified the top recruiters. Determine how long they spend on the website and thus the effectiveness of the training tools Take into account performance

	experience with data.	knowledge to make use of this data.		alterations following training, grouping based on comparable prior performances.
Can we build a better scheduling system for a hairdresser?	The manager will be able to identify prospective areas for growth by looking at the barbers' free and busy periods.	Large quantities of pre-existing data are produced by an online reservations system.	Analyzing the current data will help the barbershop understand how to use it to its fullest potential.	Identifying what employees have been successful and what employees may need retraining or refocusing.

Key themes from the United Kingdom case studies

- Existing unused data:

One of the key themes here is that many firms had data but simply had not analysed the data. Suggesting a difference between potential and realised absorptive capacity, as the firms knew to collect the data but not how to utilise it.

- Usage of Customer relationship management systems:

One of the key outcomes from these case studies is that multiple firms utilised a customer relationship management system in achieving their data project. Demonstrating the importance of identifying key software and services for enterprise to utilises in enhancing their ability to interact with data.

- Agri-tech Business:

One of the key sectors that were identified within the UK case studies was the Agri-tech sectors. Further discussion of this sector is considered within the next section.

- Interacting with high skills firms:

One of the more interesting themes which emerged was how firms with high data maturity could still benefit through identifying new potential for data.

- Online booking systems:

The enterprises utilised online booking systems which became a key source of data.

Germany

Project Question	Goal	Data	Use of Data	Business Value
How can we maximise our online presence through social media marketing to boost our client reach and sales?	Increasing contact with potential clients in order to convert them into paying customers and increase sales.	Limited internal data was available, and large amounts of an external social media database.	We've created social media workshops and examined how to enhance data from (social media) for marketing purposes.	They were able to contact more potential customers and thus increase the number of purchases.
How can we maximise our online presence through social media marketing to boost our client reach and sales?	Increasing contact with potential clients in order to convert them into paying customers and increase sales.	Limited internal data was available, however large amounts of external social media databases.	We planned a hands-on workshop for utilising data collection with Google Business.	This eventually resulted in higher growth for the business as sales rose. Additionally, they learned more about the various approaches to becoming more data mature.
Can we gain a better understanding of the untapped possibilities in our customer data by analysing and organising it?	The objective is to organise and improve internal data in order to analyse customer data.	Data was available on their customers	The data was reorganised and analysed, clearly identified key customer qualities.	They were able to be more innovative because they had greater data maturity, which gave them more insight into how

				to use the available data more effectively.
How can we set up a simple first analysis in Excel to get more information, for a company with low data maturity?	The corporation began growing its databases, but it hasn't yet gained any understanding of its information. They want to improve their data literacy so they can analyse their data in Excel.	Currently, all data is in excel.	The data usage within excel was expanded, introducing the usage of Pivot charts in examining their data.	Their perceived knowledge awareness rose as a result of the new ideas they had on how to extract more value from their data.
Is it possible to combine internal and external data in order to gather new knowledge and make data-driven decisions?	Analysing and integrating data from an electronic cash register system.	Two data sources were used: External weather data and data from their cash register.	We utilised various tools from FBD to assist them in gaining deeper insights into the data and identifying trends between the weather data and the sales data.	These insights led to an increase in perceived knowledge awareness.

Key themes from the German case studies

- Social media:

Multiple German firms consider how data could be specifically used to assist in connection with social media.

- Excel:

Some firms have data in excel, can excel be used as a vehicle for early data analysis?

- Expanding existing datasets:

In many of these examples, the firms had existing data, which was expanded upon, this is considered more in section 2.

Finland

Project Question	Goal	Data	Use of Data	Business Value
Can we introduce techniques to increase the data-driven approach of multiple firms through a generalised approach.	Encourage them to become more data-driven and provide them with a basic understanding of how to use data.	The data is from internal company datasets	The data was analysed using multiple FBD tools and through utilised Power BI.	Through this workshop, which served as their first introduction to Power BI, we motivated people. They returned home with a working project plan and a preliminary Power BI dashboard.
Determine the appropriate menu item price for restaurants.	Prior to determining the appropriate pricing for a menu item, it is vital to gain a	The data is from internal company datasets on menu prices	We have developed an Excel-based application that offers	Eventually, they gained a clearer understanding of the actual purchasing prices of their

	deeper understanding of the costs of these various menu items.	and customer outcomes	entrepreneurs with a greater understanding of the costs of various menu items.	"ingredients," including the human labour expenditures (costs). Therefore, they may now develop new menus and prices for each dish.
Increase bookings through improved customer targeting.	The proprietor of the campground had a hunch that seasonal tendencies exist for various consumer groupings. For instance, families like to book over the holidays, whereas seniors prefer to travel outside of the holidays. Thus the goal was to identify if these seasonal relationships did exist.	Seasonal data on booking information.	The SME participated in the workshop and utilised the tools provided during the various sessions. After that, they created a fresh marketing campaign based on prior bookings' insights.	The campsite owner acquired knowledge of when, how, and what types of reservations were made. There were discernible tendencies detected. The new advertising strategy increased bookings, particularly during the off-season.
Creating customised newsletters to increase sales	Target the clients more precisely and send them customised	Data from an existing client database and	The Rheezerbelt created a new	Following the distribution of customised newsletters, they

	newsletters, which contain more relevant information for them.	a newly created reservations system.	online booking system. Which reorganised its client database and created a new automated newsletter with targeted special offers.	observed an increase in the sales of particular offers.

Key themes from The Finish Case studies

- Existing unused data:

One of the key themes here is that many firms had data but simply had not analysed the data. Suggesting a difference between potential and realised absorptive capacity, as the firms knew to collect the data but not how to utilise it.

- Excel:

Some firms have data in excel, can excel be used as a vehicle for early data analysis?

Expanding existing datasets:

In many of these examples the firms had existing data, which was expanded upon, this is considered more in section 2.

- Online booking systems:

The enterprises utilised online booking systems which became a key source of data.

Belgium

Project Question	Goal	Data	Use of Data	Business Value
How can we reduce the time required to obtain management dashboard insights.	Increase the number of repeat consumers and, consequently, customer loyalty.	Excel is utilised in conjunction with available data from their ERP (Enterprise resource planning) system for administration, order processing, and planning.	The data was collected and analysed through the implementation of the ERPNext system. This system is free and an open source and provides a cloud solution.	The introduction of an ERP system not only improves their ability to target returning consumers but also enables the company to continue growing.
How can technology be used to improve job searching for individuals with Autism.	Employing someone with autism necessitates a deeper grasp of the job needs. Thus, the aim of this work is to develop a tool that identifies relevant profiles and reduces manual labour while taking into consideration all requirements.	Very limited external data was available, as such a questionnaire was utilised to identify the specific issue faced by Autisitic individuals in their job search.	Created a dedicated Excel tool to better match individual profiles to job requirements.	They increased their level of innovation by minimising the time required to make a match, hence boosting their level of production.

Can data from a CRMs be utilised to identify nearby parents and guardians in order to increase socialisation of children in the program.	Increase the number of youngsters involved in this initiative. To accomplish this, it is necessary to identify possible nearby parents or guardians and thus arrange more socialisation between children.	The information utilised is from their CRMs.	We have made modifications to the existing CRMs to improve reporting and dashboard capabilities about customer interactions and success.	They increased their efficiency by spending less time finding fresh leads.

Key themes from The Belgium case studies:

- Existing unused data:

One of the key themes here is that many firms had data but simply had not analysed the data. Suggesting a difference between potential and realised absorptive capacity, as the firms knew to collect the data but not how to utilise it.

- Usage of Customer relationship management systems:

One of the key outcomes from these case studies is that multiple firms utilised a customer relationship management system in achieving their data project. Demonstrating the importance of identifying key software and services for enterprise to utilises in enhancing their ability to interact with data.

Section 2 Discussion from themes identified in the case studies

Customer relationship management systems as a gateway to innovation and data maturity

Across multiple case studies, CRMs (Customer relationship management systems) were key in enabling the SMEs to successfully adapt and implement new processes, data collection systems and even company practices. For example, an accountancy company converted their entire client base onto an online CRMs, thus fundamentally transforming how they interacted with their clients.

Pedron et al. (2018), argue that CRMs play a pivotal role in driving innovation capability, by leading the company to develop different organization capabilities which can in some cases lead to innovation. If we consider this through the framework of *Potential* and *Realised Absorptive Capacity* that we have discussed previously within this work, CRMs can be seen as a crucial way in increasing both types of absorptive capacity. They may increase the *Realised Absorptive Capacity* by implementing digital innovation directly into their business and may increase *Potential Absorptive Capacity* due to introducing the business to an online evolutionary platform. As CRMs are forms of online platforms, the survival of the online platform is dependent upon continuous imitation and innovation, due to the ferocious nature of competition between platforms (Zhao et al., 2020). Thus, these authors propose that the innovation on the platforms can be seen to have the potential to spill over into the businesses. Therefore, CRMs may provide not only an immediate realised absorptive capacity increase but additionally may increase *Potential Absorptive Capacity* in the future, from the firm being introduced to new concepts through innovation in the platforms. One potential new type of innovation which could be crucial to innovation in CRMs over the next few years is a focus on sustainable business (Gil-Gomez et al., 2020).

Thus, CRMs is an essential tool in driving data-driven approaches in business. However, our interaction and usage of CRMs have provided us with additional insights into their specific adoption.

1) It is important to understand that each CRMs is a separate platform, each offer different features, have different requirements, and costs different amounts of money. Simply telling a business to go find a CRMs is almost the equivalent of telling someone to get a degree to increase their employment options, while this is probably true it doesn't provide them a clear answer in selecting a system which will help them. Thus, before interacting with the firm, it is important to understand the range of CRMs that are available, and this will change based on location, although CRMs are generally international there can be smaller CRMs which function at a more local and may provide features unavailable in a larger CRMs. As such the CRMs must

fit the business and do so research should be carried out on the possible choices after identifying the requirements of the firm.

2) CRMs platforms can be specialised for specific industry, such as the accountancy industry, if you can identify a CRM system within the industry then often it will have features uniquely needed for that industry. For example, when working with an accountancy firm one of the features of an accountancy-driven CRMs was to automatically look up some UK government information based on stored account numbers.

3) Simply selecting the CRMs is not enough, you also need to assist companies in transferring their clients onto the online database. Often this will only take a few hours, and in many cases, CRMs enable the import of excel sheets, however, we found that the SMEs still needed some help, especially in dealing with formatting issues. It was essential to not just tell but directly show the benefits of this tech.

There is much work to be done in identifying the specific ways in which CRMs can be utilised to help SMEs, however, these authors would argue that it is clear that they should be considered a key tool in furthering the data maturity of SMEs.

Excel as the first step in data maturity automation

One of the key questions which were posed within this project is how it is possible to carry out a data project for a firm with very limited data skills and what is possible in these circumstances. Now, the majority of the firms across all regions did have some existing level of data maturity, as shown in Figure 2 below.

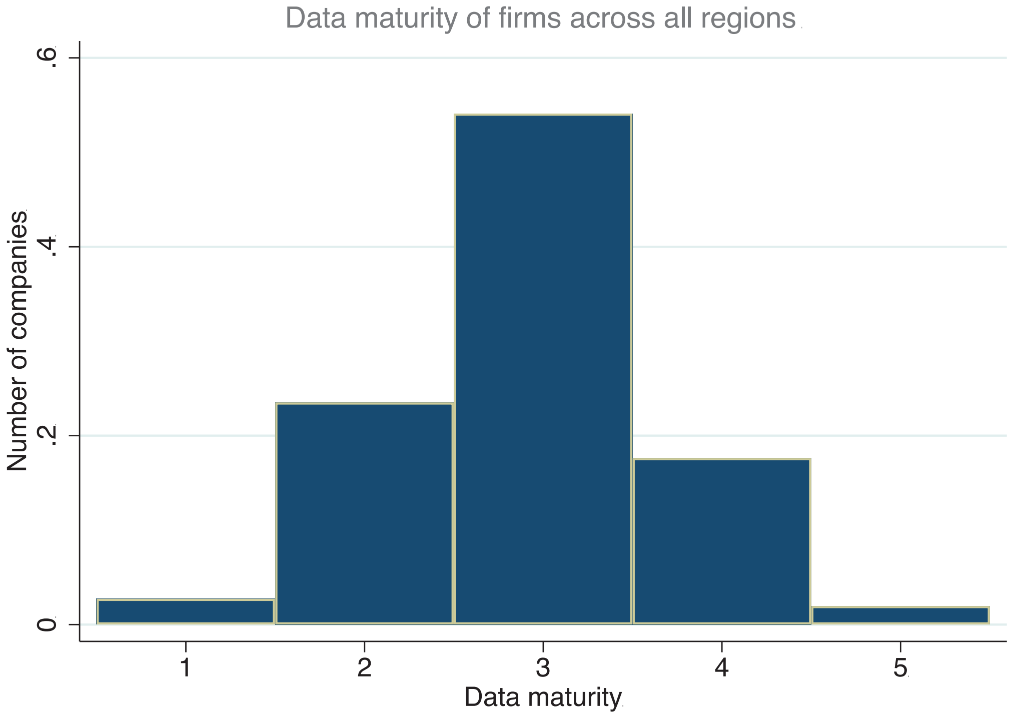

Figure 2 Data maturity of firms across all regions

However, there were still a handful of firms that had a very low level of data maturity. In these cases, we found that Excel is an excellent tool for enhancing business processes, and to start the data maturity journey. This was for the following reasons

1) Access: Almost all SMEs utilised Microsoft office, thus they already had access to excel, thus not need to install, or buy and new software.
2) Ease of progress: Some progress could be made simply by showing individuals how to do mathematical operations, in multiple cases, some owners were unaware of how to calculate the numbers in Excel and would manually use a calculator before inputting the numbers into Excel. Thus within a few minutes business processes could be revolutionised.
3) Ability to build complex models: Additionally in multiple individuals in the project built complex models within Excel, whereby all the participants had to do was input its Sales data, thus giving them access to some analysis without having to fully train the individual in excel.

4) Ability to scale: Microsoft Excel can be used for very complex business processes, utilising advanced features such as visual basic can enable even automation to be inbuilt into the software (Babkin et al., 2019).

Data maturity and type of data activity

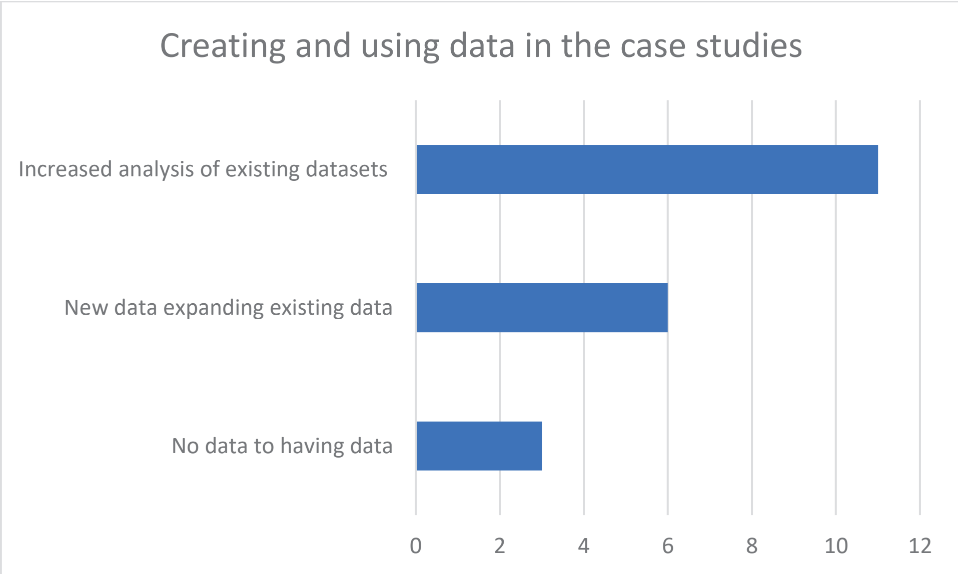

Figure 3 How the companies in the case studies have expanded their data.

The figure 3 above shows that most of the case studies had some forms of data prior to the project start, only 3 of the case studies started with no data. The other 17 had existing datasets and either expanded their data with new datasets or increased the analysis of existing data. Thus, from this one could argue that the firms we are likely to approach on this project are going to have some form of existing data which either needs to be expanded upon or enhanced. Thus, the process should not be seen as a data collection procedure but rather a process upon which untapped internal and external data sources can be used to remove barriers to innovation growth and productivity (Vasile, 2021).

Therefore, taking these points into account and our personal experiences in data discovery in relation to firm interaction the following key points were observed:

1) The first step should be to identify what data the firm has and consider that the firm owner may not recognise a specific data source. For example, many firms where using booking systems which were collecting data but were unaware, they could examine this data and utilise it in informing their business decisions.

2) It is important to identify the external data sources for the specific industry, for example one of our firms worked within the theatre sector and required data sources for this sector in a specific geographical region to support funding bids.
3) There is often one key individual who knows where the data is located within the business, this may not be the owner, it is crucial in early meetings to identify this specific member of staff (if they exist), as it they can be crucial in increasing the likelihood of a successful project.
4) When adding in new data sources it is important to consider the format of the old data source, for example if the firm original dataset is stored in excel, it can be beneficial to present new data in excel.

Agritech sector

One of the crucial sectors which we interacted with was the Agritech industries, the authors considered that this sector interacted with data in a very unique way and thus have provided Agritech specific case studies. Before considering the key points derived from these

Agritech Case study 1: Antobot

Antobot is a robotics technology business that develops high-tech practical robotic solutions for numerous horticultural skills, including crop yield status, crop yield prediction, pest and disease detection crop management, and precision farming support. To achieve this, they have created autonomous robots that can be deployed directly on farms, utilising cutting-edge technology. Contrary to the majority of companies in this industry, their product is already in use on farms and is actively transforming farmers' understanding of their crops and land.

The Antobot system is not merely a standalone robot, but rather a fully integrated data collecting system that utilises and expands upon existing agricultural technologies to enable farmers to understand and improve their farms while remaining compliant with existing and emerging regulatory problems. Thus, in considering the digital and technical skills of the company, this company can be seen to have the highest levels of technical and digital skills, as demonstrated by their ability to create automated machinery.

The primary objective of the Future by design project was to explore what information they would desire to acquire regarding their customers' data. Currently, they were collecting data on each client, but it was viewed as more of a bureaucratic procedure than one that could aid in the company's growth. The FBD team members analysed the employee's history and skill sets before recommending that data collection from the companies be approached from a machine learning standpoint. In other words, what variables would best help the organisation to make future decisions about what work has been effective and what has not, and what variables do they need to collect to assess their business success over time? Considering that variables may always be removed if they are not useful for determining or comprehending the business's success, new variables can only be added in the future, as a broad variety of variables must be collected. Thus, it was suggested that a specific customer relationship management system be built to ensure that a defined set of variables for each customer is captured. Permitting them to utilise their extensive technical talents in analysing the successes and failures of their company.

Agritech Case Study 2: Potato Consultancy

The examined consultant is a world-renowned expert in their profession. Prior to becoming a private consultant, he worked as a researcher at a research centre. Some of his services included providing farmers with data for irrigation scheduling, while others were more conventional agronomic services centred on production methods. The major of focus for the SME was the move from a typical work atmosphere to consulting. While the consultant had a grasp of the data requirements and structuring of customer data around this, the systems are unsophisticated and wasteful, making use of spreadsheets for example. The consultant feared that abandoning a known method would hinder his capacity to evaluate the quality of his services. The clientele he had and his goals for the business were appropriate. However, organising of client contact details through a customer relationship management system was considered as a way of improving his business processes.

Agritech Case Study 3: Aponic Aeroponic Farming Systems

Aponic Aeroponic Farming Systems is a vertical farming solutions company aiming to revolutionise the commercial and individual farming industry through building aeroponic vertical farms using nutrient solution in replacement of soil and have been found to dramatically reduce water usage while increasing crop yields. Additionally, crops can be harvested continuously throughout the year reducing instability in output in comparison to traditional farming methods, furthermore this can transform the commercial structure of the farm, enabling them a more stable and continuous income. The companies additionally offer specific training to ensure that farmers are staying ahead in this ever-changing environment, through its membership options.

In this specific case the focus for Futures by Design was more of a consultancy focus. Focusing on considering what specific areas of expansion are a possibility for this growing business. Utilising the current data, they are collecting to consider what factors may be affecting the company over the coming years. For example, we considered whether additional staff could be necessary and how the owner would automate production and installation of the systems.

Agritech Case study 4: Sylvia Newman Limited:

Sylvia Newman are a garden and design company that for over 20 years have been producing a wide range of garden design services both for commercial and non

The focus on the future by design project in this specific case was examining how stability in the management of data could be improved, through examining the current structure of data which is summarised Figure 4. Thus, the question was posed what would happen to the data if the single key employee was unwell or left the company? Would the company be able to accurately find the data, or would the data even be lost? Thus, the first step suggested regarding the company was to consider a system for backing up this core data.

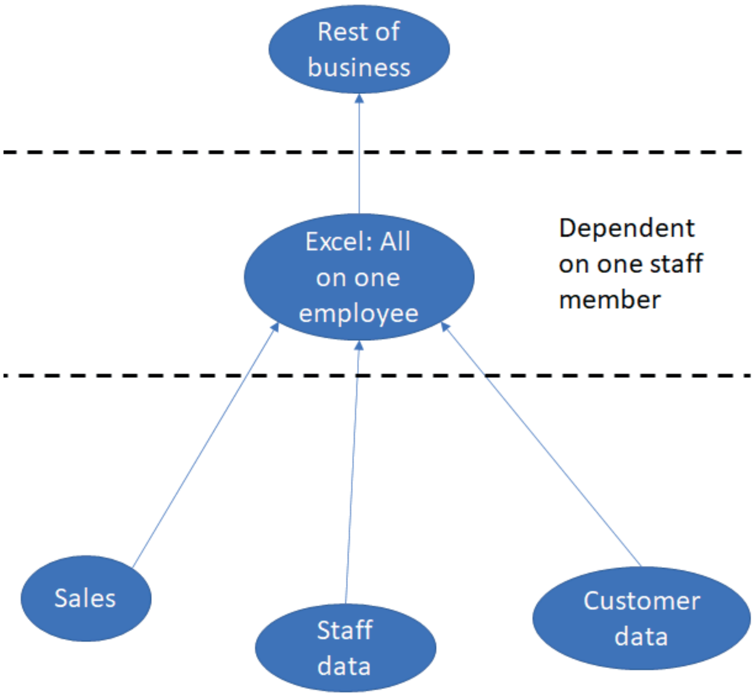

Figure 4 Diagram representing the current structure of data within the company

Summary points from the Agritech sector

Thus, the following key points were derived from these case studies and connected Agritech data project:

1) Data is often abundant with farmers often having far too much, some farmers had literally entire sheds full of data.
2) Food regulations and requirements are key factors in driving data innovation, there is also fears that even more requirements are going to be implemented in the future.
3) Emerging new agri-technology companies are data-driven; however, they may not use their data in support of the business aspects of the work.

Section 3 Quantitative analysis: The impact of Futures by Design

In this section of the report, we focus on a quantitative analysis of the impact of *Futures by Design*. Given the data availability, the key direct empirical evidence of the impact for our programme, to be used for this quantitative analysis, is the evidence about data projects' completion by the participating Small and Medium Sized Enterprises (SMEs).

The development of a firm's specific data project was one of the most relevant steps in our digitization activities with SMEs. Indeed, FBD has developed a set of tools helping to outline the current situation of a participating SME, define the key problems it faces, and explore the possibilities for improvement. Before starting a data project, these tools are useful to know the data maturity level of an organisation. In this part of the report, we will explore the emerging statistical relationships between the key features emerging from different tools capturing some of the key features of the participating organisations and the outcome variable specifying whether the participant has successfully completed the chosen data project within the FBD timeframe.

The key features we focus on, are those capturing digital competencies, attitudes, together with different aspects of a participant organization's capacity to absorb, explore and assimilate internal and external sources of digital knowledge. These were collected through the *Data Jumpstart Questionnaire*. This tool, developed by the Team of Futures by Design, is a Questionnaire consisting of a set of 40 questions that dive deeper into various aspects of data maturity (The Questions are reported in Appendix 1).

The *Data Jumpstart Questionnaire* explores through set of interrelated questions, the digital: infrastructure, tools and culture characterising an organization prior to start working on a dedicated data project with our team.

Every company that has embarked on an the FBD process has completed the *Data Jumpstart Questionnaire*, as one of the first steps to identify their digital needs and barriers. After filling the answers into the *Data Jumpstart Questionnaire*, a dedicated team of FBD researchers identified, together with the SMEs, a tailored data project aimed at the adoption/introduction of a digital innovation to address the SME's identified priorities and needs.

Our Key Outcome Variable: Project Completion

The first part of this report discussed some of the key qualitative evidence emerging from the analysis of the specific projects. In this second part, instead, we focus on quantifying the effect of the key drivers, captured from the *Data Jumpstart Questionnaire*, in determining an SME's likelihood to succeed, or fail, in completing the data project, i.e., to adopt a digital innovation.

Ideally, one would like to measure our impact on SMEs with the detailed financial data on an organisation turnover before and after the implementation of the data projects.

However, the data available on the *proportion of turnover due innovation*, collected through the *Data Jumpstart Questionnaire*, only refer to the period preceding the FBD's project intervention. For this reason, our focus is on the only observable outcome of our intervention based on the close interaction with an SME: its projects' completion. Moreover, circa 95% of companies reported that, as the result of their project completion, they increased their productivity and ad growth, through adopting the data project Innovations. Hence, by choosing projects' completion as our outcome variable in the quantitative analysis, we obtain an informative proxy, on SMEs' overall impact on their growth, innovation, and productivity, due to our interventions.

Futures by Design (FBD) supported SMEs from six North Sea European regions to innovate, grow and increase productivity by making better use of data and their digital skills. As discussed above, the key activity of FBD was to identify and to implement a digitization project for each participating SME. By the time of writing, FBD has brought to successful completions 92 data project, out of 256 participating SMEs. These projects enable participating SMEs to make a major step towards being better equipped for competing in the digital age, increasing the perspectives for their economic survival and market success.

A *Project completion* is, in itself, a kind of digital innovation introducing different ways of improving businesses activities through the adoption of data solutions and digital technologies. Indeed, FBD data projects lead SMEs to introduce products/ and or service updates or innovations, leading to increased growth and productivity. Therefore, the outcome, in terms of success of failure, of project completion provides an indicator for the success of FBD projects and an opportunity for informing other SMEs and stakeholders about how they could use their data to increase their business successes.

To capture our key outcome variable, *Project Completion*, we construct a dummy variable indicating whether a company has brought a data project to completion, within the agreed FBD timeframes. These data are collected across the six North Sea regions through a dedicated dataset the "*Sqans dataset*", constructed, managed, and updated by the FBD Team.

We consider completion as the most reliable indicator of FBD project outcomes, as we are not yet able to see future turnover impact, due to innovations.

Given the choice of project completion as our main outcome variable, the key question we want to address with our quantitative analysis is "*What are the underlying factors determining the probability that a firm will complete a FBD project?*".

To address this question, we will focus on the information obtained from the subset of participating SMEs that completed a pre-project *Data Jumpstart Questionnaire*. We will then use the information contained in these questionnaires to explore different effects of key variables on the likelihood of an SME completing their data project.

Descriptive Analysis of the Project Completion

To describe the key features of data projects, we start by focusing on the distribution of their completions across sectors and regions. These data will provide us with an initial idea of potential differences across our six North Sea regions and sectors, as potential factors influencing data project completions. Sectorial and regional evidence of the evidence on data project completion is summarised in the two key figures below (Figure 5).

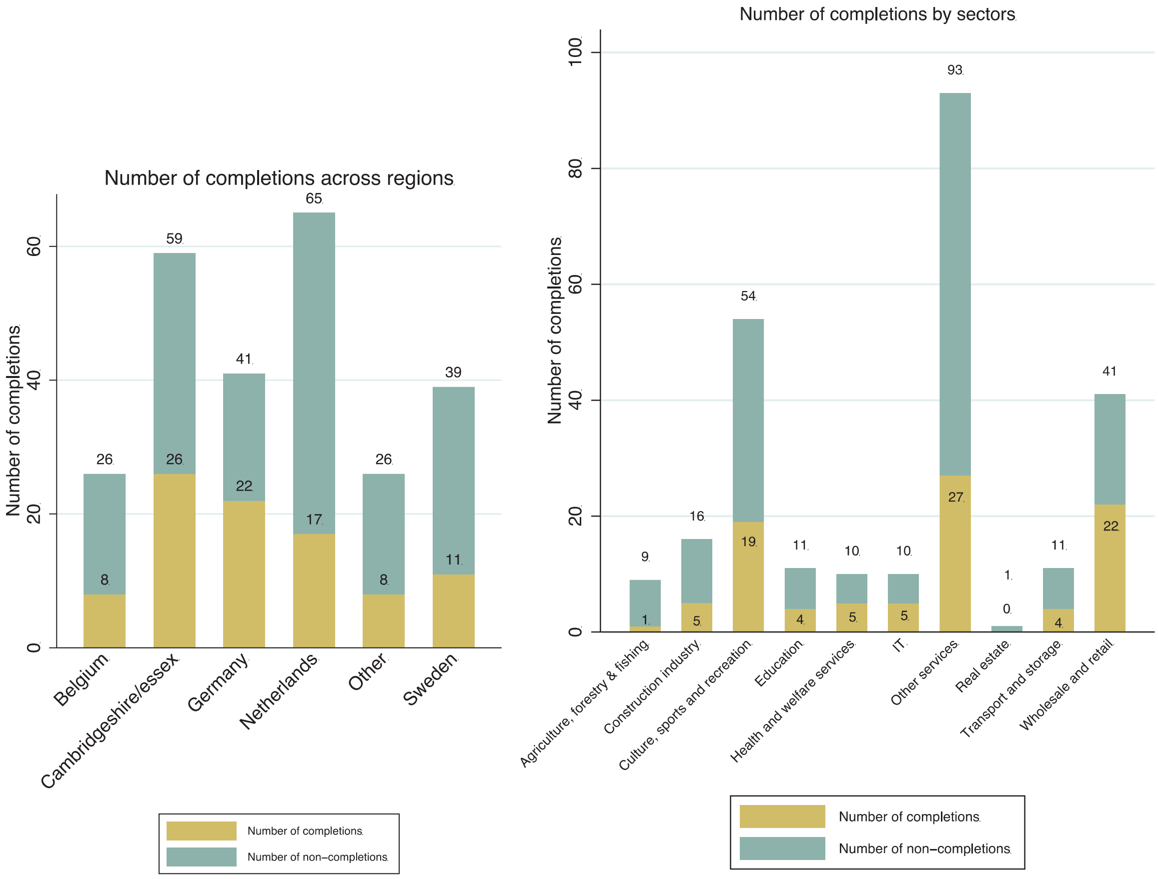

Figure 5 Number of completions across regions and sectors

Looking at the number of FBD projects, independently of their completion, the first evidence we notice is that of a great variation across sectors and regions. The highest number of projects, 65 projects, was started in Netherlands, followed by the United Kingdom's 59 projects and Germany's 41 projects. However, interestingly, Germany and United Kingdom's have a comparably high completion rate, 54% and 44% respectively. Whereas the Netherlands reported the lowest completion rate of 26%, among the six of North Sea regions.

Form this distribution we can see that FBD has helped our six regions, focussing on an early stage of digitalization to identify the potential for digital upskilling, by working on data project improving existing/development of new products, organisational methods, and markets. Furthermore, FBD projects seem to mostly accelerate the pace of digitalization of wholesale and retail, Culture, sports and recreation and other services, with 54, 41 and 93 projects respectively. In turn, the figures reflect the need of growth through digital innovation in these sectors. While sectors like IT, Healthcare and welfare services and wholesale and retail sectors, given higher observed data maturity levels, are more likely to complete the digital transition and unleash their businesses growth.

The determinants of project completion

To identify the key drivers and barriers of project completion for SMEs, we start by focussing on two different aspects of SMEs abilities related to their data cultures and attitudes. Following Zahra and George (2002), these can be classified as:

- *Potential Absorptive Capacity* (PAC), which includes the capacities of an organization to **acquire and assimilate** relevant, data-based, knowledge, and
- *Realized absorptive capacity* (RAC), which, instead, focuses on an organization capacity to **transform and exploit** such, data-based, knowledge.

Figure 6, below, provides a breakdown of these two different aspects of the original concept of absorptive capacity (Cohen and Levinthal, 1990). This division will enable us to clearly identify which SME characteristics are responsible for the companies' ability to *acquire* and *assimilate* knowledge from external sources and which characteristics enable the companies to process (transform) and apply (exploit) this external digital knowledge.

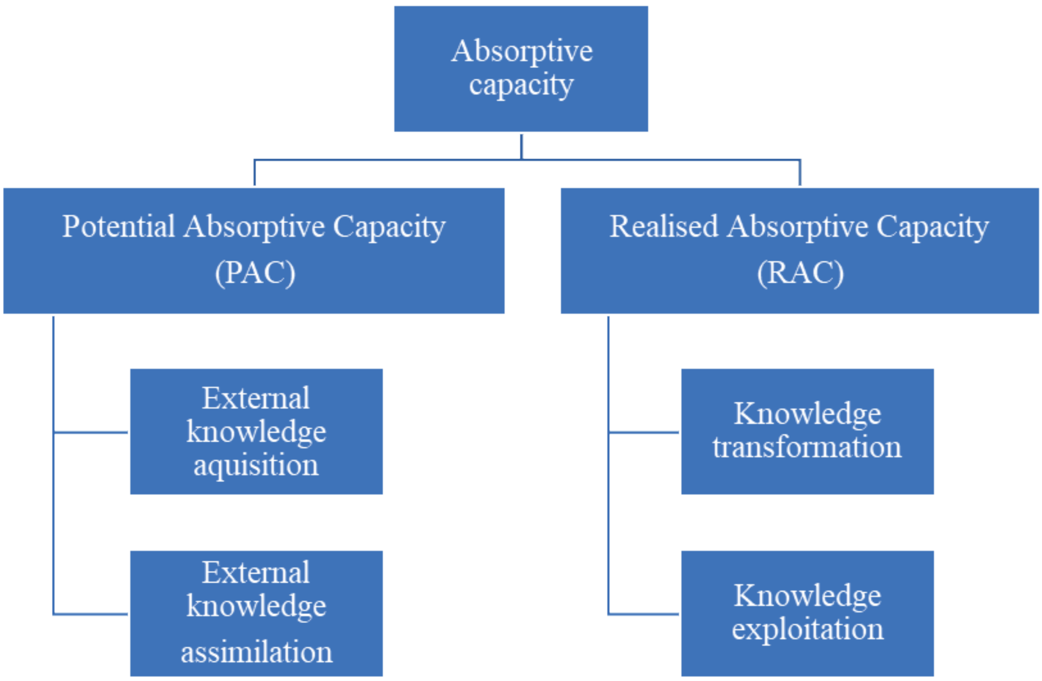

Figure 6 The breakdown of absorptive capability

Operationalizing PAC and RAC

In order to operationalise these two concepts, capturing different aspects of digital absorptive capacity, and then utilising to ascertain their role and impact on the likelihood of an SME completing its dedicated FBD data project, we follow a conceptual framework developed from the contributions of a large body of literature (Jansen et al., 2005; Zahra and George, 2002; Szulansky, 1996).These authors' validation work led us into identifying the relevant questions to focus on within the *Data Jumpstart Questionnaires*, and to map them into the PAC and RAC constructs.

In detail, we identified four separate *Jumpstart Questionnaires* items to measure the intensity of the *Potential Absorptive Capacity* as the efforts made by an SME in *acquiring* and *assimilating* new external knowledge pertaining to data and digital innovations. Similarly, *Realised Absorptive Capacity* was measured through a total of five items from the *Data Jumpstart Questionnaires*, to assess the extent to which firms are able to *transform* and *exploit* external digital and technological knowledge, i.e., to use it towards introducing innovations.

In Table 1, below, we provide the detailed mapping between the specific questions from our *Jumpstart Questionnaire* and the key components of PAC and RAC identified in the aforementioned literature. The scores for the items are then obtained from a 5-point disagree-agree scale obtained on the responses of the survey participants in the *Data Jumpstart Questionnaire*.

Table 1: Mapping questionnaire questions with PAC and RAC

No.	Items	Matched Jumpstart Questionnaire Answers
***Potential Absorptive Capacity* (PAC)**		
1	New opportunities to serve our clients are understood rapidly	*Q19.1 My colleagues often bring new ideas and developments with regard to data to the table*
2	We analyze and interpret changing market demands promptly	*Q19.3 My organization strives for fast adoption of innovations in the field of data*

3	Employees record and store newly acquired knowledge for future reference	*Q20.1 My Organization is aware of the possibilities of working with data*
4	We quickly recognize the usefulness of new external knowledge to existing knowledge	*Q20.4 My organization often takes part in events with data as one of the main topics*
Realized absorptive capacity (RAC)		
1	We incorporate external technological knowledge into our firm	*Q20.2 My organization likes to work with external parties when it comes to data gathering and analyses*
2	We thoroughly grasp the opportunities new external knowledge offers our company	*Q20.6 When new data becomes available, I use this to review my opinion*
3	We periodically meet to discuss the consequences of market trends and new product development	*Q19.2 My colleagues in general know their way around with new data-related technologies*
4	Employees are clearly aware of how the firm's innovation activities should be performed	*Q19.6 I am confident that the data within my organization is up to date*
5	We are constantly reviewing how to better exploit external knowledge	*Q19.4 When it comes to data, my organization has the means and opportunities to implement new developments quickly*

After mapping our *questionnaire* answers into the above items, identified in the current literature as indicating PAC and RAC, our next step in the quantitative analysis of the drivers of project completion, required to build two operational constructs, with a clear semantic interpretability. These two constructs were created by performing a *Principal Component Analysis* on the different items related to the dimensions of *Realised* and *Potential Absorptive Capacity*, reported in Table 1 above.

In detail, the method used, *Principal Component Analysis* (PCA), provides a data reduction technique aimed at decreasing the number of variables in an analysis by describing a sequence of uncorrelated linear combinations of these variables that account for the majority of the data variance. Additionally, along this process of dimensionality reduction, the key information from a PCA can be examined to learn more about the underlying data structure and hence to obtain a semantic interpretation of the achieved decomposition (For an introduction, see Rabe-Hesketh and Everitt (2007, chap. 14)).

Principal component Analysis

PCA seeks to identify linear unit-length (L0L = I) combinations of the variables with the highest variance. The first principal component has the greatest variance overall. The second principal component has the greatest variance among linear combinations of unit length that are uncorrelated with the first principal component, etc. The variance of the final principal component is the smallest among all linear combinations of unit-length variables. All principal components contain the same information as the original variables, but the important information is distributed differently across the components: the components are orthogonal, and earlier components contain more information than later components. Consequently, PCA is merely a linear transformation of the data. It does not assume that the data comply with a particular statistical model, but it does require that the data be interval-level data. (See Stata, 2015)

Figure 7 Principal Component Analysis

In our analysis, we loaded only the factors that were included as proposed indicators for each item as represented in Table 1. The logic is that only factors that explain at least the same amount of variance as a single variable is worth keeping. Hence, Statistically, we always select the factors with an eigenvalue of ≥1, as which accounts for as much variance as a single variable. Hence, we use the scree plot to help us with the factor selection. Figure 7 shows the eigenvalues on the y-axis and the number of factors on the x-axis. The point where the slope of the curve is clearly levelling off (the “elbow) indicates the number of factors that should be generated by the analysis. The results of the PCA show that a single factor model fits the data moderately well and was therefore used as Proxy for PAC. Similar results were obtained for RAC (see Figure 7 for both RAC and PAC). Once these constructs for PAC and RAC were obtained, these recombined versions of PAC and RAC, were utilised as the key explanatory variables in the analysis of the success or failure of project completions.

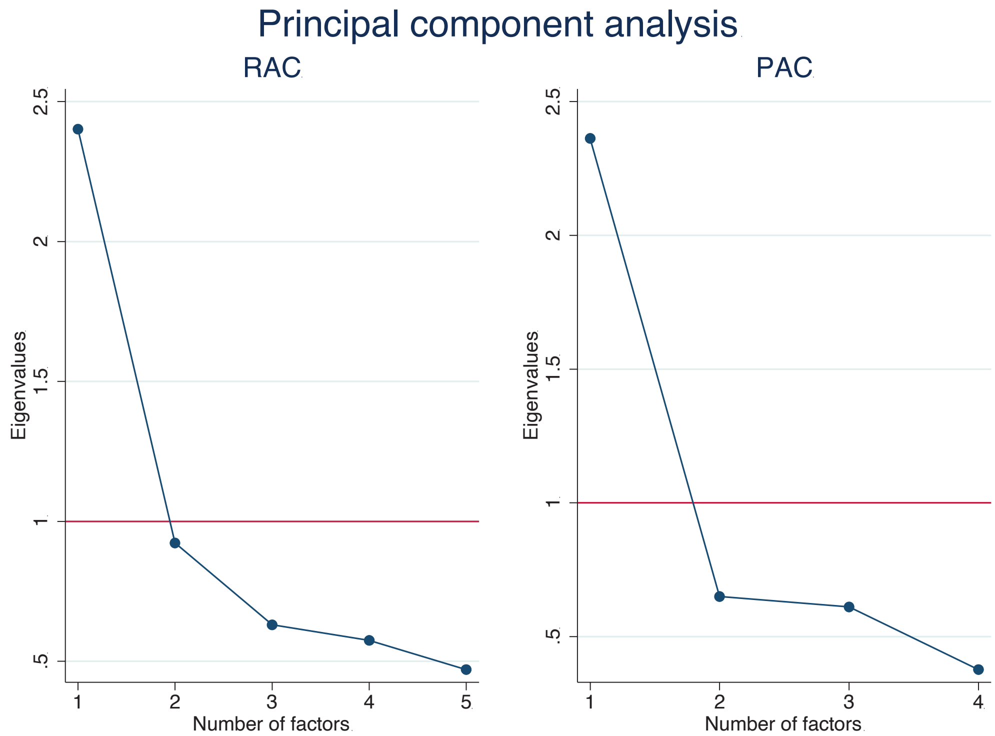

Figure 8 Graphs of PCA Thresholds

Additional Control Variables

In addition to PAC and RAC, this report also considers the role of other key characteristics relating to dimensions that can affect an SME's dedicated FBD data Project completion's outcome. This is essential, both for reaching more coherent estimates of the impact of PAC and RAC, and, also, to obtain a better understanding of how multiple drivers compound or cascade innovation and digitalization for SMEs. Table 2, below, summarizes the specific alternatives answers associated to each one of the additional *Jumpstart Questionnaire* questions investigated and used as additional control variables in our estimates of the probability of an SME successfully completing a dedicated FBD Data Project.

Table 2: Categories of the control variables used in modelling FBD Data Project completions

Potential determinants	Components						
Internationalisation: Target Market	Local	Regional	National	EU Wide	Worldwide		
Awareness about digital tools	Spreadsheets	SQL	Visualisation	Application program interface (API)	Reinforcement learning		
Type of Data Availability	Product Data	Financial Data	Customer Data	Employee Data			
Experience of having previously introduced an Innovation for:	Producing goods	Logistics	Supporting	Organizing			
Time allocation across data processes	Gathering Data	Managing Data	Analyzing Data	Visualizing Data			
Security and compliance	Overall data security importance	Software	Consistency	Regulatory awareness	Regulation compliance		
Type of data Infrastructure	Data warehouse	Spreadsheets	Cloud	External	Visualization	Statistical	Programming
Data Quality	A synthetic measure of the overall SME data quality.						
FTE	The number of full-time employees in a company						
Company age	The number of years since a company was founded						

Modelling strategy

In this section, we build a series of regression models using an Instrumental Variable Probit regression (IVPR) approach to assess, separately, the individual effects of both PAC and RAC on the likelihood of an SME's dedicated FBD data project completion.

We also consider the additional effects on project completion exerted by other key control variables capturing different relevant dimensions of our SME's. These control variables focus on key concepts, captured by the *Jumpstart Questionnaire* addressing the degree of an SME's *Internationalisation, Awareness about digital tools, Type of data availability, Experience of having previously introduced innovations, Time allocation across data processes, Awareness about Security and compliance, Type of data Infrastructure* and additional controls such as *Company's size*, captured by the number of employees, and *Company age*.

We opted to use an IVPR model, since this modelling approach is deemed to be appropriate when there is a plausible reason to believe that one or more of the explanatory variables are endogenous, i.e., they are correlated not only with the dependent variable, as hypothesised, but also with other possible variables not included in the model but still affecting the dependent variables (Omitted Variables being part of the error term). Hence, an IVPR fits models with binary dependent variables and covariates that might be affected by endogeneity.

Instrumental Variable Probit regression

Formally, the model is

$$y_{1i}^{*} = y_{2i}\beta + x_{1i}\gamma + \mu_i$$

$$y_{2i}^{*} = x_{1i}II_1 + x_{2i}II_2 + \nu_i$$

Where i=1,…,N, in our case the number of SMEs in our sample. y_{2i} is a $1\times p$ vector of endogenous variables, in our case, PAC and RAC, x_{1i} is a $1\times k_1$ vector of exogenous variables, in our case capturing the additional SMEs dimensions discussed above, x_{2i} is a $1\times k_2$ vector of additional instruments, in our case capturing the company's sector, and the equation for y_{2i} is written in reduced form. By assumption, the two error terms for the two equations are multivariate normally distributed, with mean zero, and Variance Covariance matrix Σ : $(\mu_i, \nu_i) \sim N(0, \Sigma)$, where the variance of the firs term σ_{11} is normalized to one to identify the model. In the first equation, β and γ are vectors of structural parameters, capturing the effects of the endogenous and the exogenous variables, respectively. In the second equation, estimating the endogenous variables y_{2i}^{*}, II_1 and II_2 are matrices of reduced-form parameters for the other exogenous variables, x_{1i}, and for the instruments, x_{2i} . The IVPR is a recursive model: y_{2i} appears in the equation for y_{1i}^{*}, but y_{1i}^{*} does not appear in the equation for y_{2i}, hence the estimated values for y_{2i} are not correlated with the error term μ_i addressing the original endogeneity problems.

Regarding the binary nature of the dependent variable, project completion, the idea is that we do not observe, the continuous latent variables leading to project completion, y_{1i}^{*}; instead, we only observe whether a project was completed

$$y_{1i} = \begin{cases} 0 & y_{1i}^{*} < 0 \\ 1 & y_{1i}^{*} \geq 0 \end{cases}$$

The order condition for identification of the structural parameters, β and γ, requires that the number, k_2, of instruments, x_{2i}, be higher than the number, p, of endogenous variables, y_{2i}, hence: $k_2 \geq p$. The Variance and Covariance Matrix, Σ, is expected not to be block diagonal between μ_i and ν_i, otherwise, y_{2i} would not be endogenous variables, as there would be no correlation between the endogenous variables and the error terms of the original fist stage equations, no omitted viable bias.

Figure 9 Instrumental Variable Probit regression

Hence, in our estimates with the IVPR model, we follow a recursive analysis: we use Probit regression of SME's FBD dedicated data project's completion, with exogenous variables describing the degree of internationalisation, awareness of digital tools, availability of data (data per business segments), time allocation across data processes, awareness about security, infrastructure, company age and size (i.e., the number of employees). However, we also consider, in this stage, the endogenous variables, capturing our constructs for PAC and RAC, that are separately recursively estimated, using additional instruments based on regional dummies to reduce /eliminate their potential endogeneity.

Empirical Results

The primary purpose of this report is to identify the main determinants of the completion of an FBD data project by looking at key different aspects and organizational characteristics of the SMEs that took part in the FBD programme. These features are a firm's absorptive capability (we considered separately both the PAC and RAC), having innovation experience, internationalization linked to the target market, the level of awareness about digital tools, the awareness about data security, the availability of data (data per business segments), the time allocation across different types of data processes, the data infrastructure, company age and size (i.e., the number of employees). The empirical results, obtained through our IVPR models are discussed in the following.

Absorptive capability (PAC and RAC) and FBD Data Project Completion

Innovation activities often involve multiple capabilities, based on the ability of searching, processing, and integrating knowledge. Innovations are therefore the result of processes involving the use of newly created knowledge that a firm has been able to absorb. Absorptive Capacity (Cohen and Levinthal, 1990) is therefore the key factor that influences the likelihood of a company's innovation (Yu, 2013). Firms with greater absorptive capabilities tend to enhance their learning capabilities, which helps them effectively utilize external knowledge. In more detail, the success in implementing an innovation depends both: on a firm's capacity to acquire and assimilate, i.e., on a company's *Potential Absorptive Capacity* (PAC), (Leal-Rodríguez et al., 2014) and on its ability to transform and exploit external knowledge, i.e., on its *Realized absorptive capacity* (RAC). As discussed in the introduction, we follow Zahra and George (2002) and decompose absorptive capacity into its two these different constructs, to study their effects on the likelihood of FBD Projects' completions.

This separation between PAC and RAC enables us to develop clear insights into how these different perspectives, realised or potential, of a company's absorptive capability may play potentially contrasting roles on the underlying incentives and technological capabilities (Von Tunzelmann, 2009) of an organisation leading to the adoption of data driven innovations, captured in our model by the successful completion of a tailored data project.

Figure 8 below, depicts these effects, as estimated through our two IVPRs. The full set of estimates is reported in the Appendix.

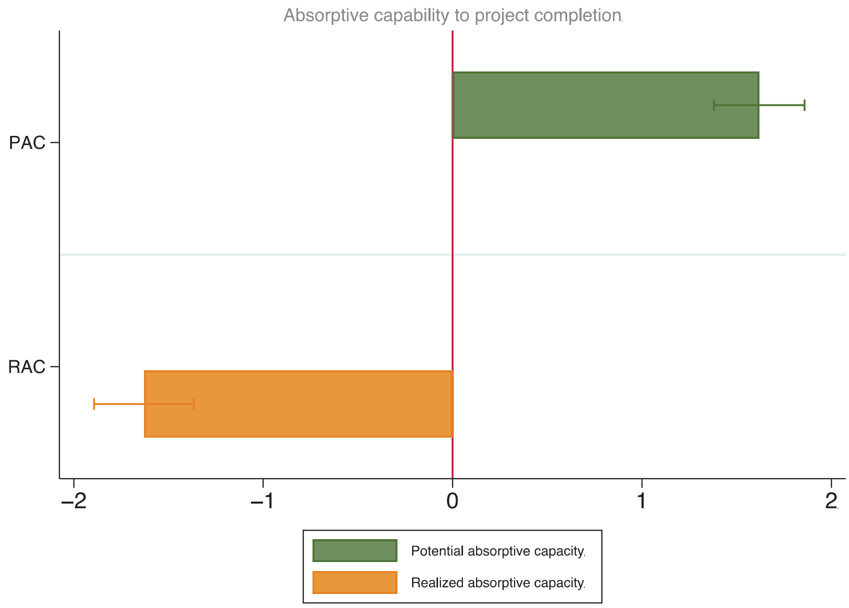

Figure 10 Potential and Realised Absorptive capacity and the likelihood to complete an FBD data project

The empirical evidence collected by FBD and analysed through two separate IVPR models: one controlling for the PAC and the other for the RAC, shows that while the PAC of an organization displays a significant and positive impact (1.619***) on the likelihood of a project completion, its level of RAC shows, on the contrary, a significant but negative impact (-1.63***) on this likelihood.

Hence, our findings clearly indicate that: while a firm's ability to acquire and assimilate external knowledge, its *Potential Absorptive Capacity*, or dynamic capability (Winter, 2003), facilitates the completion a FBD data project, (a process of adoption of data driven innovations), the opposite effects is exerted by a company's *Realised absorptive capacity*, holding back the likelihoods of a company's project's completion.

This dichotomy seems to indicate that when an organisation is more focused on the set of its *technological competencies* (Iammarino et al., 2012) by *transforming* and *exploiting* knowledge that has been already achieved, i.e., its *Realised absorptive capacity*, this organization becomes less likely to embark into, or to bring to a fruitful conclusion of, a new data innovation project. When on the contrary, such an organization focuses on its abilities to *acquire* and *assimilate* external technological knowledge, i.e., on its *Potential Absorptive Capacity*, it becomes more likely to successfully complete the data projects, by fully capturing its *technological capabilities*.

Internationalization and FBD Data Project Completion

Most SMEs from our sample trade within their local market with only a few businesses operating at an EU-wide or worldwide market level. Therefore, firms rely heavily on the knowledge sourced targeting markets locally, regionally, and nationally. Those firms trading nationally have a relatively higher exposure to external resources and markets hence, they can source from different types of technology, resources, and knowledge, that can be acquired, absorbed, and utilized by these firms.

According to findings from *Recombinatory search theory* (Savino et al., 2017), for a firm to innovate, it requires the ability to recombine its current knowledge, problems and solutions, all activities for which knowledge exchange is important (Fleming and Sorenson, 2001). This is the case since, an organisation's knowledge that is relevant for innovation is often tacit and embedded, which means that it cannot be codified and hence obtained through contractual market exchanges (Nelson and Winter, 1982). In this sense, knowledge transfers within local markets are likely to be more efficient, since such interactions/collaborations reduce the ambiguity of knowledge that a firm obtains from other firms, thus facilitating knowledge transfer and assimilation (Jensen and Szulanski, 2007). Knowledge transfers within a nation or even globally, instead, might reduce the likelihood and efficiency of knowledge transfers and delay innovations, or in our case a project's completion. The extent of internationalisation of a firm is often captured in terms of whether the firm sells products/services in Regional, National, EU or Rest of the World markets. These variables are widely used in the innovation literature to control for the impact that global competition exerts on innovations (See, for example, Zoia et al., 2018; Archibugi and Iammarino, 1999). Our *Jumpstart Questionnaire* addressed the role of internationalization with a question, Q7, asking: *In which geographic markets does your enterprise sell goods and/ or services? Local, Regional, National, EU-wide* and *Beyond EU.*

In Figure 9 below, we report the impact on completion of the degree of internationalisation, based on the target market: separately for the two models, one based, including RAC and the other including PAC. Then for each one of these models, we also see how these different target market affect PAC and RAC. The full results from our two IVPR models are reported in the Appendix.

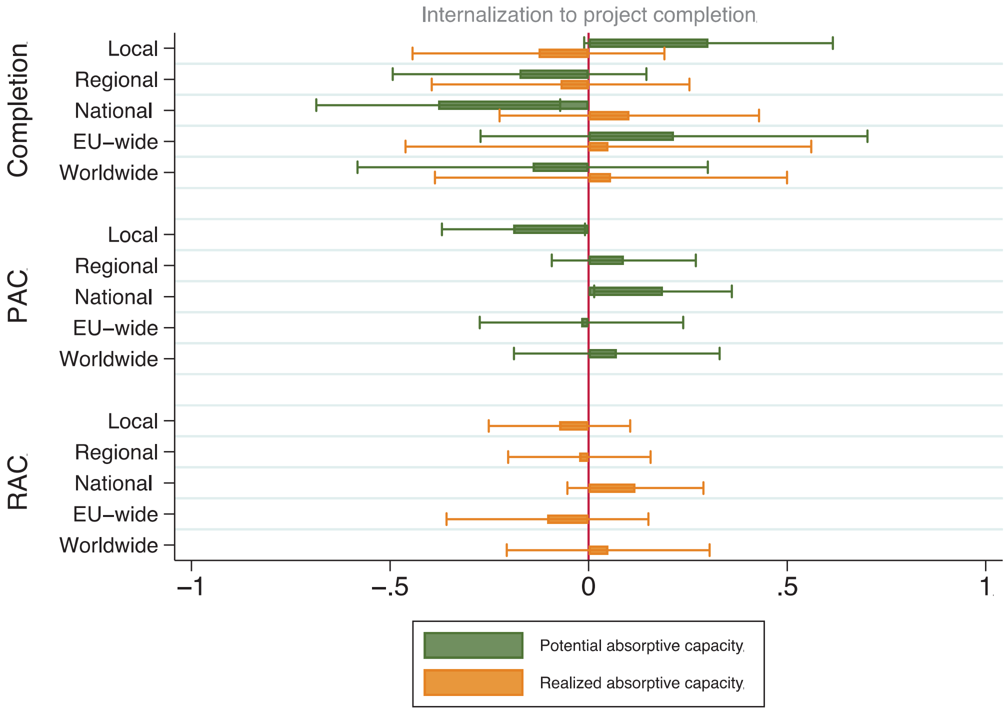

Figure 11 Impact on completion of the degree of internationalisation

The empirical evidence indicates significant findings on the effects of internationalization on project completion only appear in the IVPR model where we control for the PAC. In this model, indeed, the coefficient describing the effects on project completion, when an SME answers that it is trading in *Local Markets,* is positive and significant at 10% (.302*); while targeting *National Markets* has a negative and significant at 5% effect (-.379**) on the likelihood of the completion of FBD data projects when conditioning for PAC, capturing the ability of a company for acquisition and assimilation of new external knowledge.

Internationalization does not appear to exert any significant impact when conditioning for RAC, the ability for knowledge's *transformation* and *exploitation*. This is as it would be expected since *Local market targeting might be culturally easier than even National market Targeting,* as a framework for completing FBD data projects, the nature of which was often highly local. These effects disappear when controlling for the *Realised Absorptive capacity*, that is clearly capturing already existing complementarities, but it is important when controlling only for the *Potential Absorptive Capacity* that is, on the contrary, focusing on external knowledge assimilation, that might indeed be facilitated by trading at the local market level.

Awareness of Digital Tools and FBD Data Project Completion

The progress of an organization's digitalization clearly depends on its level of digital awareness. SMEs that are aware of the advantages of digital tools, such as reinforcement learning, frequently favour greater technological integration and list the adoption of new technologies among their top goals (Garzoni et al., 2020). As a result, these organisations acquire the abilities needed to respond successfully and swiftly to market changes and achieve long-term economic expansion.

In our case, the *Jumpstart Questionnaire* contained a question, Q15, about the company's awareness about digital tools. It asked: *To what extent have you heard about the following tools and techniques:*

- *15.1 Spreadsheets*
- *15.2 Structured Query Language (SQL)*
- *15.3 Data visualization tools (Tableau, Kibana, PowerBI)*
- *15.4 Application Programming Interface*
- *15.5 Reinforcement Learning*

In Figure 10 below, we report the impact on completion of the Awareness of digital tools, separately for the two models, one based, including RAC and the other including PAC. Then for each one of these models, we also see how Awareness of digital tools affect PAC and RAC. The full results from our two IVPR models are again reported in the Appendix.

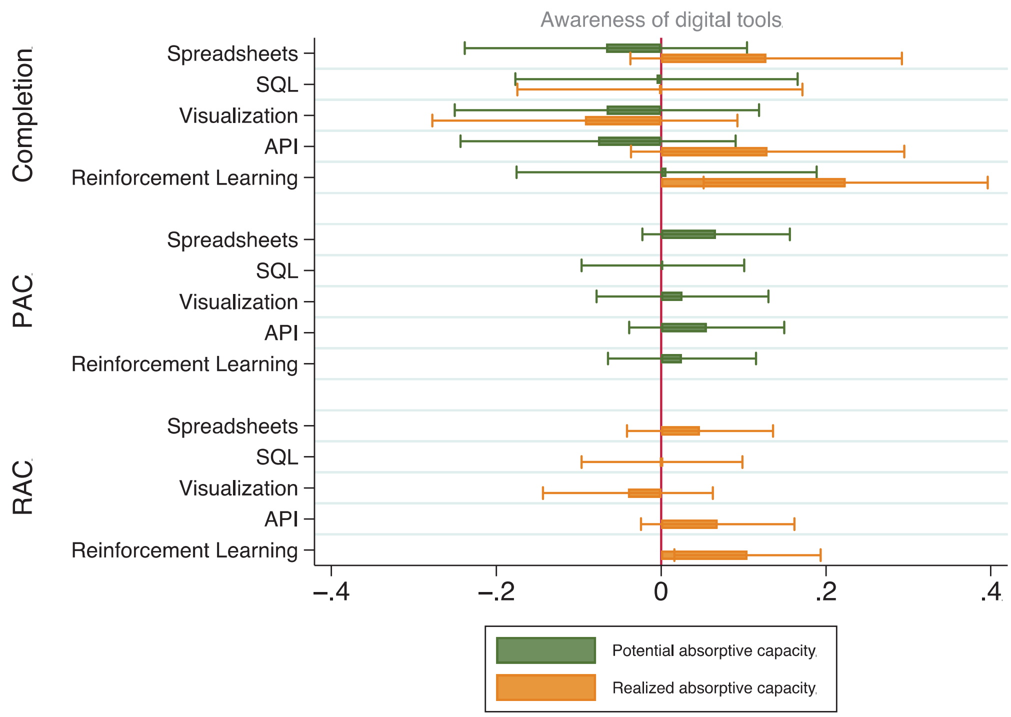

Figure 12 Awareness of digital tools to project completion

Interestingly, from Figure 10 above and the estimates in the Appendix, we can see that in our IVPR models there is a positive and significant at 5% effect (.224**) of the awareness about the tools and techniques of *Reinforcement Learning* on the likelihood of the completion of FBD data projects when conditioning for the RAC, capturing the ability for knowledge's stead transformation and exploitation. No significant effect was instead found when conditioning on the PAC, capturing the ability for knowledge's acquisition and assimilation.

Data Availability and FBD Data Project Completion

SMEs are often afflicted by poor efficiency of data usage due to their limited financial and human resources. Furthermore, Data availability is expected to play a critical facilitating role the implementation of data innovation processes. However, the shifting focus across different available data might also introduce a trade-off by reducing the attention and focus on applying new knowledge to innovate and grow. In our questionnaires we dealt with Data Availability and Data Processing times, through question, Q16, on data availability asking: *My organization analyses the following data:*

- *16.1 Product data (e.g., sales)*

- *16.2 Financial data (e.g., billing)*
- *16.3 Customer data (e.g., orders, contracts)*
- *16.4 Employee data (e.g., absence. Productivity)*

In Figure 11 below, we report the impact on completion of the Data availability, separately for the two models, one based, including RAC and the other including PAC. Then for each one of these models, we also see how Data availability affects PAC and RAC. The full results from our two IVPR models are again reported in the Appendix.

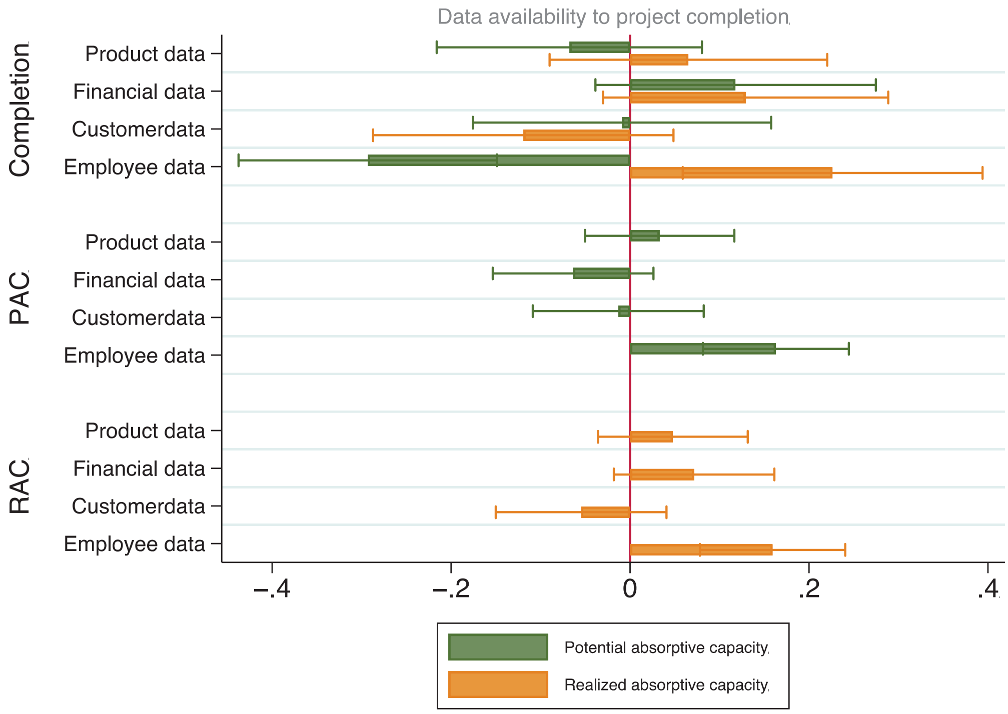

Figure 13 Data availability and project completion

The only significant effects of Type of data availability on data project completion are those related with the availability of employee data. In detail, the analysis of employee data, shows contrasting and significant effects at 1% (-.293***) for IVPR model controlling for PAC while an opposite positive and significant at 1% effect (.226***) for the model controlling for RAC. The negative effect seen in the PAC model; seems to reflect a tendency to a shift of the focus from seeking out external knowledge to internal operations when we control for acquisition and assimilation through the role of *Potential Absorptive Capacity*.

When, on the contrary, we control for the *transformation* and *exploitation* dimensions of absorptive capacity, i.e., for the *Realised Absorptive Capacity*, the impact of *Employee data* is positive and significant, as, in this context, *Employee data* become a key element required to track the success of innovations. The analysis of *Employee data*, e.g., absence or productivity, reflects how a business executes its core processes and delivers value to its customers and informs the executive team as to whether their data projects are working.

Time allocation across data and FBD Data Project Completion

Time spent on gathering or digitizing data provides wider opportunities for business operations to improve accuracy and analytical power for decision-making process. On this dimension, the *Jumpstart Questionnaire* contained a question, Q18, asking "*Within my organization we spend a lot of time on":*

- *18.1 Gathering and digitizing data*
- *18.2 Managing and maintaining a database / datafiles*
- *18.3 Running data analyses (e.g., comparative statistical analyses, predictive analyses)*
- *18.4 Visualizing data (e.g., making graphs)*

In Figure 12, below, we report the impact on completion of the Time allocation across data processes, separately for the two models, one based, including RAC and the other including PAC. Then for each one of these models, we also see how Time allocation across data processes affects PAC and RAC. The full results from our two IVPR models are again reported in the Appendix.

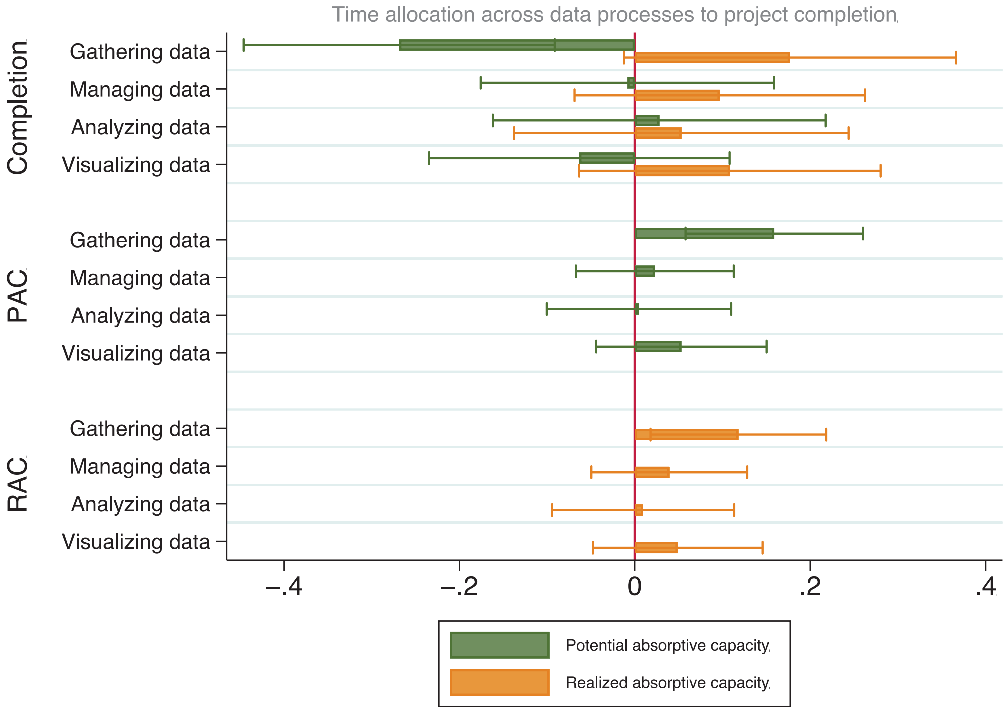

Figure 14 Time allocation across data processes and project completion

The key finding in this case points to different effects on project completion. There is a negative and significant at 1% effect (-.269***) of the time allocated to *Gathering and digitizing data* on the likelihood of the completion of FBD data projects when conditioning for PAC, capturing the ability of a company for acquisition and assimilation of new external knowledge, while, the same time allocation activities have a positive and significant (at 10%) impact (.177*), when conditioning for RAC, the ability for knowledge's transformation and exploitation. These contrasting effects are as it would be expected since *Gathering and digitizing data* is expected to complement positively with the *Realized Absorptive Capacity*, while seems to be superfluous, and hence negative, when considered with the *Potential Absorptive Capacity*.

Previous Innovations and FBD Data Project Completion

The innovation literature has benefited from extensive work using data regularly collected by Eurostat within the *Community Innovation Surveys* [1] to understand the key drivers underlying different types of innovations. The classification of innovations derived from *OSLO Manual* (OECD, 2005), distinguishes between: *Product innovations* as "new or significantly improved goods or services," *Process innovations*, defined as "new or significantly improved methods for the production or supply of goods or services" and *Organisational innovations*, defined as "new business practises for organising procedures, new methods for organising work responsibilities and decision making, or new methods for organising external relationships with other firms or public entities."

Iammarino et al. (2012) and Zoia et al. (2018) chose the affirmative responses to questions on having introduced *product* and *process innovation*s as indicating businesses having *technological capabilities*. Similarly, these authors viewed the negative responses to these same questions, matched with the presence of investment in innovative activities, as indicating firms with *technological competencies* but lacking the capacity to transform their *competencies* into innovations.

In this report, we follow in the spirit of Iammarino et al. (2012) and Zoia et al. (2018) and use the *Jumpstart Questionnaire* answers on previously introduced innovations as proxies for our companies' *technological capabilities* (Von Tunzelman, 2009) and estimate the impact of these *technological capabilities*, on the likelihood of completion of an FBD data project.

In more detail, the *Jumpstart Questionnaire* contained a question, Q14, asking: "Have you introduced or improved any of the below in the last 2 years?

- 14.1 Methods of manufacturing for producing goods or services?
- 14.2 Logistics, delivery or distribution methods for your inputs, goods or services?
- 14.3 Supporting activities for your processes, such as maintenance systems or operations for purchasing, accounting or computing?

[1] See https://ec.europa.eu/eurostat/web/microdata/community-innovation-survey

- 14.4 New business practices for organizing procedures (i.e. first time use of supply-chain-management, business re-engineering, knowledge management, lean production, quality management, etc.) using data-informed decision making?

In Figure 13, below, we report the impact on FBD project completion of an SMEs' *Technological capabilities*, captured by having previously introduced any of the above innovations. This is done separately for the two models, one based, including RAC and the other including PAC. Then for each one of these models, we also see how *Technological capabilities* affects PAC and RAC. The full results from our two IVPR models are again reported in the Appendix

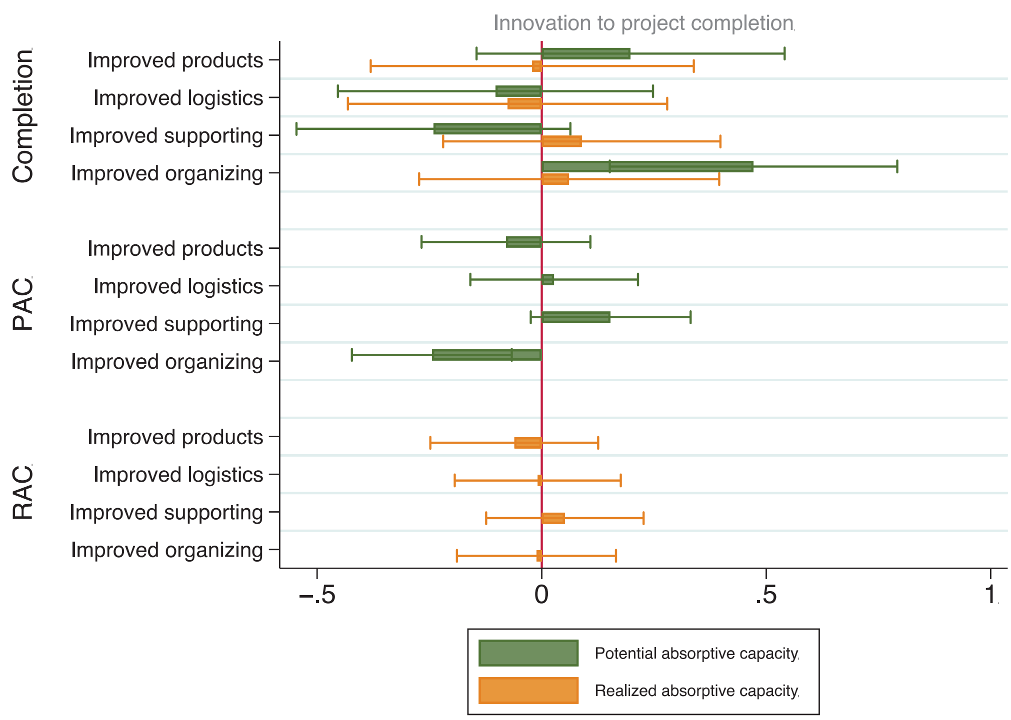

Figure 15 Previous innovations and FBD project completion

From Figure 13, above, and the estimates tables in the Appendix, one can see that only the introduction of "New business practices for organizing procedures (i.e. first time use of supply-chain-management, business re-engineering, knowledge management, lean production, quality management, etc.) using data-informed decision making" helps with the completion of FBD projects having a positive and significant (at 1%) effect (.472***), when conditioning for PAC, capturing the ability of a company for acquisition and assimilation of new external knowledge.

These results might indicate that the relevance of our intervention, was implicitly more successful for companies likely to have *technological competencies, rather than capabilities, since there are no positive significant effects of having in the past introduced neither product nor process innovations*.

Digital Infrastructures and FBD Data Project Completion

In an economic environment increasingly characterised by the relevance of big data and by their analysis through deep learning techniques, the use of these data and methods has become progressively more relevant in facilitating the process of knowledge discovery and therefore for the introduction of innovations. In detail, the recent dramatic progresses in mathematical programming (Grossmann, 2012), coupled with advances in machine learning (Jordan and Mitchell, 2015), especially in deep learning over the past decade (LeCun et al., 2015), sparked increasing interest in data-driven optimization (Bertsimas et al., 2018; Bertsimas and Thiele, 2006).

In this framework, firms having the relevant infrastructure to harnesses the increasingly rich information underlying big data can benefit from utilising smart and data-driven processes in their decision making. On the other hand, as firms acquire and assimilate knowledge based on algorithm and experience, they may become less open-minded to the appreciation of new external knowledge, or less in need to be assisted with tailored FBD data projects.

To further explore this issue, our *Jumpstart Questionnaire* included a question, Q17, asking:" My organization has the following digital infrastructure":

- 17.1 A central storage for all data (a data warehouse)
- 17.2 Spreadsheet software such as Microsoft Excel, Libreoffice, OpenOffice
- 17.3 Access to a cloud computing platform (AWS, Google, Azure)
- 17.4 Access to external data via APIs or scraping
- 17.5 Databases such as SQL Server or Oracle
- 17.6 Statistical software such as SPSS or Stata

- 17.7 Data visualization software such as Tableau, PowerBI or MapInfo
- 17.8 Programming languages such as R or Python

In Figure 14, below, we report the impact on FBD project completion of an SMEs' *Digital infrastructures*, captured by having any of the above types of digital infrastructures. This is done separately for the two models, one based, including RAC and the other including PAC. Then for each one of these models, we also see how *Digital infrastructures* affect PAC and RAC. The full results from our two IVPR models are again reported in the Appendix.

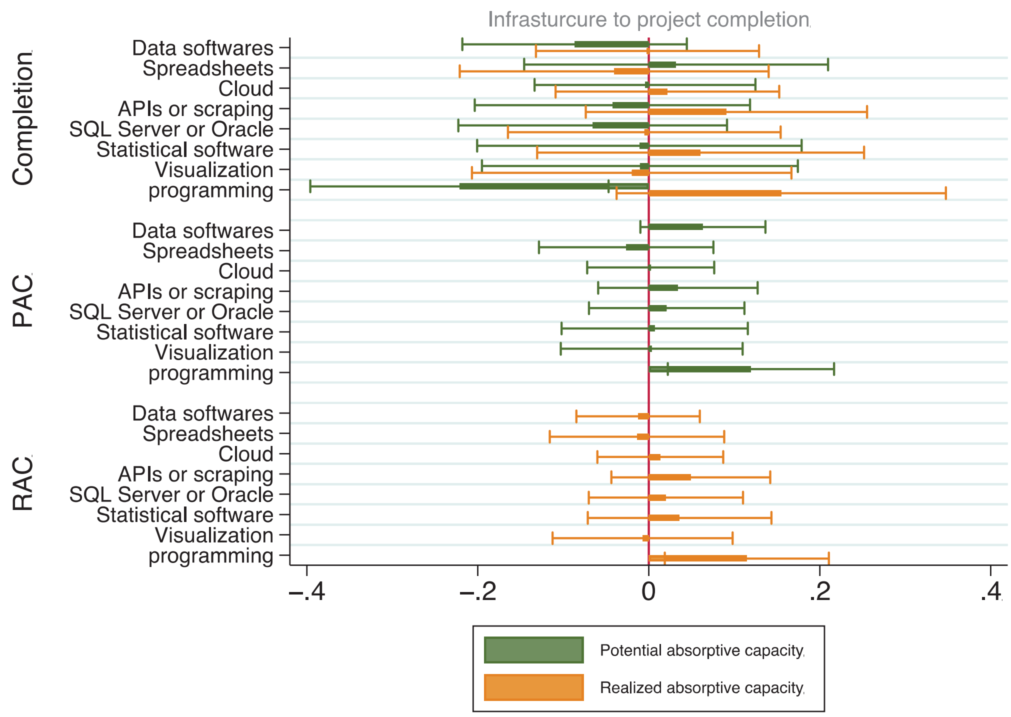

Figure 16 Digital infrastructures and FBD data project completion

The empirical evidence derived from our two IVPR models, summarised in Fig 14 above, indicates that the only negative and significant (at 5%) effect is observed when companies have data infrastructure for "Programming languages such as R or Python". This variable has in facts a negative impact of (-.222**) on FBD data project completion when conditioning for PAC, capturing the ability of a company for acquisition and assimilation of new external knowledge.

Firm size and FBD Data Project Completion

Firm size, captured by the log of firms' employment, was also used in our two IVPR models to control for the *Schumpeterian* notion that large firms are more likely to embark and succeed in innovations (Schumpeter 1911, Pellegrino and Piva, 2020 and Breschi et al., 2000). In more detail, the *Jumpstart Questionnaire* contained a question, Q9, asking: "How many FTE work in your organization?" We use this question to obtain the data of the number of full-time employees (FTE) as an indicator of firm size. On average, SMEs have 20 FTE employees, with the maximum of 500.

In Figure 15, below, we report the impact on FBD project completion of an SMEs' *Firm size*. This is done separately for the two models, one based, including RAC and the other including PAC. Then for each one of these models, we also see how *Firm size* affects PAC and RAC. The full results from our two IVPR models are again reported in the Appendix

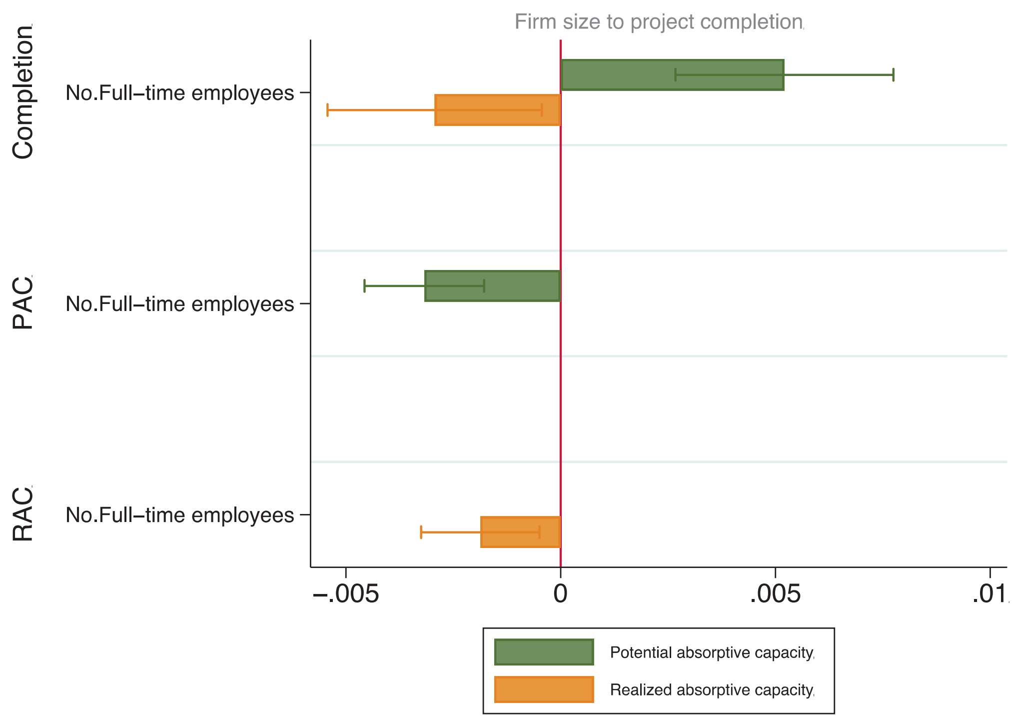

Figure 17 Firm size and FBD project completion

From Figure 15, above, one can see that Firm size has a positive and significant (at 1%) effect (.005***) on the completion of FBD data projects when conditioning for PAC capturing the ability of a company for *acquisition* and *assimilation* of new external knowledge, while is has, on the contrary, a negative and significant (at 5%) impact (-.003**) when conditioning for RAC, the ability for knowledge's transformation and exploitation. According to resource-based view, firms with more employees are more likely to bring diverse ideas, perspectives and resources that can fuel the innovation. Nevertheless, from a different information decision-making perspective, it can be argued, that more information may cause conflicts that end up reducing the efficiency of teamwork and decision making on innovation processes. Hence, from this different perspective firm size has a negative impact on innovation processes that delays the completion of FBD projects.

Sectors and FBD data Project Completion

Moncada-Paterno'-Castello (2022) and Iammarino et al. (2012) highlighted the importance of controlling for the sectoral distribution on innovation activities. In our two IVPR models we also control for the sector of a company's activity. This is captured in the *Data Jumpstart Questionnaire* surveying which branches/industry that SMEs belong to. Overall, our SMEs are from 10 different sectors, including agriculture, forestry and fishing, construction, education, health and welfare services, culture, sports and recreation, wholesale and retail, transport and storage and other services, culture, sports and recreation, wholesale, and retail. Based on this information, we could explore whether the completion of FBD projects varied across these sectors in a significant way. The empirical results showed no significant effects of the sectors, when controlling for PAC capturing the acquisition and assimilation dimensions of absorbing capacity, whereas some variations in construction, health and welfare services, IT, transport and storage and wholesale and retail sectors, are significant when controlling for the RAC, capturing transformation and exploitation stage. The empirical evidence from the IVPR model, conditioning for RAC, indicates therefore that FBD projects for companies, in these sectors, are more likely to be successfully completed, so that adoption of data driven innovations in such sectors are more likely. These industries are data-driven and quite mature in data and infrastructures, which facilitates the completion.

In Figure 16, below, we report the impact on FBD project completion of an SMEs' *Sector*. This is done separately for the two models, one based, including RAC and the other including PAC. Then for each one of these models, we also see how *the Sectors* affect PAC and RAC. The full results from our two IVPR models are again reported in the Appendix.

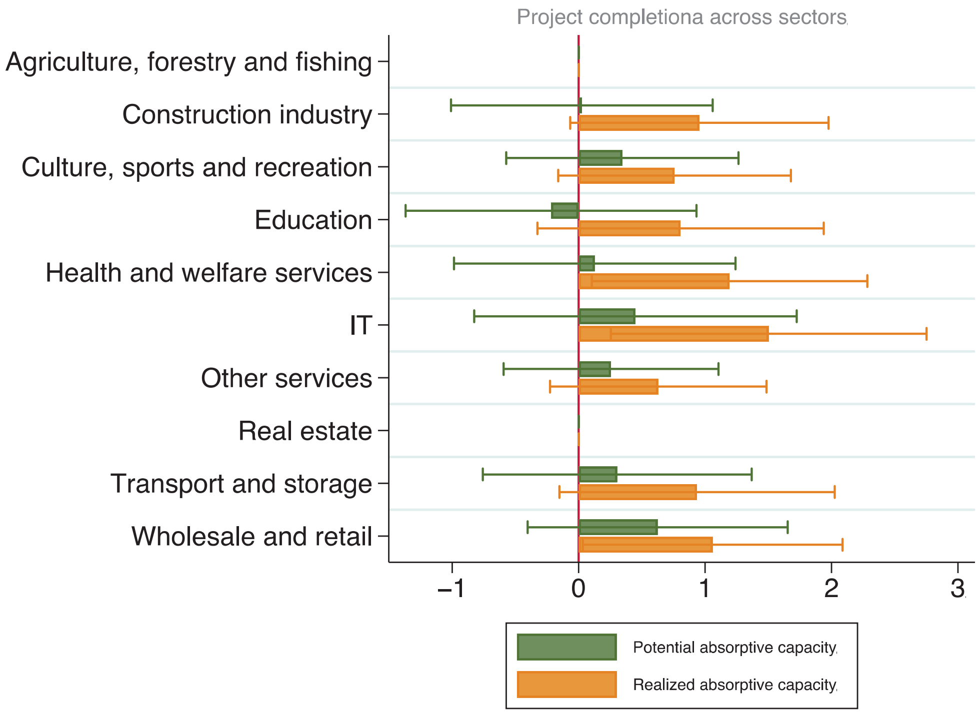

Figure 18 Sectors and FBD Data project completion

Security and FBD Data Project Completion

In the framework of data driven innovations based on data generated by users; businesses have great opportunities at creating strategies aimed at boosting their profitability. These opportunities are explored when discussing the proposed data projects with the SMEs. However, new strategies and methods of data collection to support, for example, new digital marketing strategies, may pose serious questions about user privacy (Arya et al., 2019; Roberts, 2015; Saura et al., 2021 Zuboff, 2015). These issues become more relevant, due to the new ways businesses learn from the data that customers create in digital markets, increasing the prominence of dealing with privacy problems (Schoen et al., 2013). We addressed companies' awareness and concerns about data security, in the *Jumpstart*

Questionnaire, through a dedicated question, Q.21 asking about the extend a SME would agree with the following statements:

- 21.1 My organization is aware of the importance of data security
- 21.2 Within my organization everybody uses the same software (Office, Salesforce, Dynamics etc.)
- 21.3 My organization is aware of the law and regulations when it comes to data (e.g., the GDPR)
- 21.4 I am always strictly law abiding with the laws and regulations regarding data.

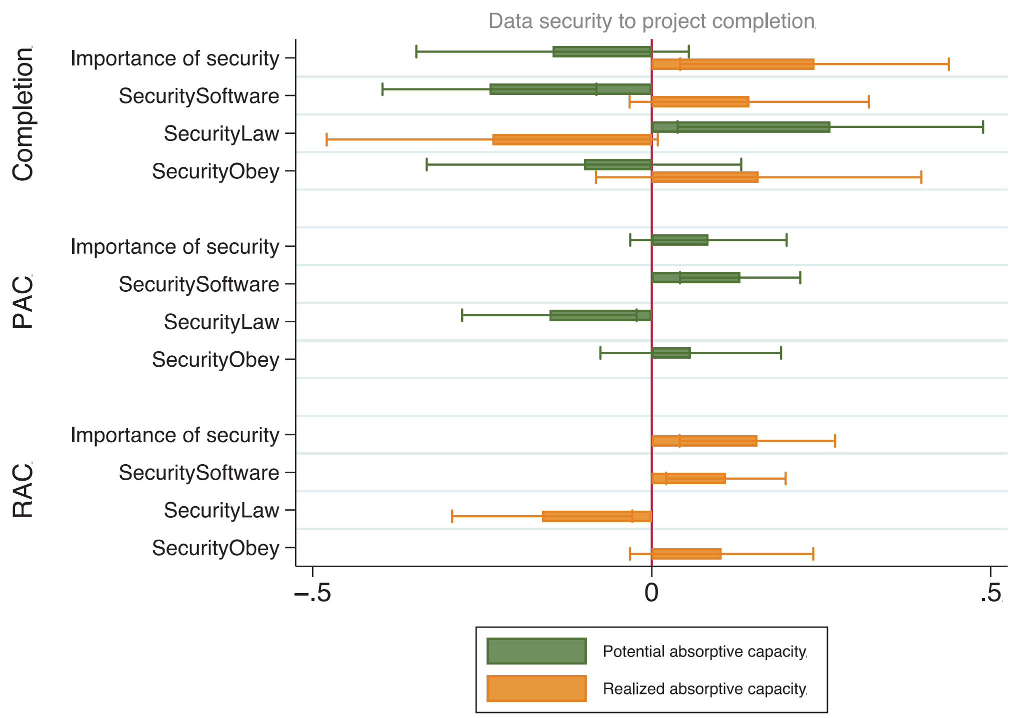

Figure 17 Data security and FBD Data project completion

The results (Figure 17) show that, in the model controlling for *Realised absorptive capacity*, an organization awareness about data security has a positive and significant (at 5%) impact (.24**) on the likelihood of an FBD data project completion, while the specific awareness of the data law and regulations, e.g., GDPR, seems to have a negative and significant (at 10%) impact (-.235*) on completion, hence raising barriers from moving from the status quo. From this perspective, awareness of the data law and regulations may come to represent the equivalent of an additional cost of adoption for data-driven innovations. In the other IVPR

Model, controlling for *Potential Absorptive Capacity*, however, the effect of legal awareness is reversed, (.263**) since legal awareness has a positive effect on FBD data project completion. In this model, the variable expressing software “my organization everybody uses the same software” capturing conformity expresses a negative and significant (at 1%) impact (-.239***) on FBD data project completion. This effect is likely to capture a cautious behaviour, highlighting the role of conformity in constraining innovations.

Conclusions & Recommendations

In this Report, we investigated the impact of FBD data-projects in six North Sea European regions. We focussed on understanding and characterising the drivers and barriers for project completion, based on the data collected through the *Jumpstart Questionnaires*. Our exciting and novel empirical evidence, collected throughout direct interaction during the Covid-19 pandemic, shows the presence of interesting variation among SMEs’ attitudes towards data usage and in their digital skills, all forming critical components of an SME’s overall absorptive capacity. Such variability is key in shaping the individual FBD data projects trajectories, and their probabilities of reaching the completion (or non-completion) with the ensuing adoption (or not-adoption) of a digital innovation. Our aim was to focus on how possible gaps in digital competencies might influence SMEs’ innovation outcomes, captured as their data-project completion, and to identify the resulting opportunities for policy interventions that could unleash SME growth potential, where it is currently lower, due to lower levels of digital skills, attitudes, infrastructures, and *digital capabilities*.

Based on our initial results, our empirical evidence allows us to derive the following recommendations

Recommendation 1. *“The success of an SME’s data-project, leading to adoption of data innovation, benefits from an increased level of SMEs’ Potential Absorptive Capacity. Hence programmes & policies should focus on training and supporting SMEs’ in Acquiring and Assimilating, external data-based knowledge.”*

Such policies, when matched with specific interventions, such as those provided by the Futures by Design tailored data projects, will increase the impact of these intervention, increasing the probability of successful data project completion and leading to increased rates of adoption of digital innovations, improved productivity and overall better, realised, or potential growth rates.

Our empirical evidence and model results also showed that high levels of *Realised Absorptive Capacity*, are a predictor for project non-completion. Hence, we propose

Recommendation 2. *"The success of an SME's data-project, leading to the adoption of a data innovation, suffers from an increased level of SMEs' Realised Absorptive Capacity. Hence programmes & policies should focus on supporting SMEs' who do not yet have high competencies and abilities to Transform and Exploit data-based knowledge."*

Such policies, within a wider plan of specific interventions, such as those provided by the FBD tailored *data projects, will increase the impact of these intervention, by avoiding wasteful intervention where its value for money (in terms of successful data project realisation) is lower due to the lesser probability of successful project completion when SMEs already have a higher ability to Transform and Exploit data-based knowledge.*

Finally, we focused on the analysis of the specific effects on the likelihood of project completion of the many covariates, capturing SME's degree of Internationalization, Awareness about digital tools, Type of data availability, Experience of having previously introduced innovations, Time allocation across data processes, Awareness about Security and compliance, Type of data Infrastructure, Company's size, and Company age. The overall recommendation emerging from this empirical evidence is:

Recommendation 3. *"A one size fits all" approach in supporting digitalization processes cannot apply across SMEs. Digital Sector Policies should tailor their incentives towards processes based on a detailed analysis of the SMEs specific characteristics in terms of their degree of Internationalization, Awareness about digital tools, Type of data availability, Experience of having previously introduced innovations, Time allocation across data processes, Awareness about Security and compliance, Type of data Infrastructure, Company's size, and Company age. Moreover, it is of paramount relevance, for these policies to be mindful of both SMEs' sectoral and regional specificities*."

This last recommendation stems from the evidence showing that companies that trade locally or regionally have preferential access to external resources, knowledge and networks that enhances the ability to transform their capabilities into innovations. Furthermore, a firm's ability of absorbing and assimilating external knowledge, captured by potential absorptive capability inspires innovative input and facilitates the completion of FBD projects. Therefore, an additional recommendation would be to extend information exchange across firms in their native network to increase their exposure to external resources. This might be achieved through the creation of networks and hubs that may facilitate collaboration between companies. Our findings support the relevance of the role of networking in driving innovation suggesting that exchanging knowledge with suppliers, customers, and intermediaries (professional and trade associations) will contribute to companies getting access to new markets and technologies (Pittaway's et al., 2004; Brunswicker and Vanhaverbeke, 2014; Dubouloz et al., 2021). Thus, governments could support SMEs' innovativeness and originality by planning for and contributing to the development of data exchange hubs and Networking trade associations aimed at fostering business collaborations (See, for example the Digital Sector Strategy for Cambridgeshire & Peterborough, 2019)

References

- Archibugi D. & Iammarino S., (1999). The policy implications of the globalisation of innovation, Research Policy 28, 317–336
- Arya, V., Sethi, D., & Paul, J. (2019). Does digital footprint act as a digital asset?–Enhancing brand experience through remarketing. *International Journal of Information Management*, *49*, 142-156.
- Babkin, A.V., Kuzmina, S.N., Oplesnina, A.V. and Kozlov, A.V. (2019), September. Selection of tools of automation of business processes of a manufacturing enterprise. In *2019 International Conference" Quality Management, Transport and Information Security, Information Technologies"(IT&QM&IS)* (pp. 226-229). IEEE.
- Bertsimas, D., Gupta, V., & Kallus, N. (2018). Data-driven robust optimization. *Mathematical Programming*, *167*(2), 235-292.

- Bertsimas, D., & Thiele, A. (2006). A robust optimization approach to inventory theory. *Operations research*, *54*(1), 150-168.
- Breschi S., Malerba F. and Orsenigo L. (2000). Technological regimes and Schumpeterian patterns of innovation, Economic Journal 110, 388–410.
- Brunswicker, S., & Vanhaverbeke, W. (2015). Open innovation in small and medium-sized enterprises (SMEs): External knowledge sourcing strategies and internal organizational facilitators. Journal of small business management, 53(4), 1241-1263.
- Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative science quarterly*, 128-152.
- Dosi, G., & Nelson, R. (2018). Technological advance as an evolutionary process. In Nelson, Richard R., Giovanni Dosi, Constance E. Helfat, Andreas Pyka, Pier Paolo Saviotti, Keun Lee, Sidney G. Winter, Kurt Dopfer, and Franco Malerba. "Modern evolutionary economics: An overview." (2018).p. 35-84. Cambridge University Press.
- Dubouloz, S., Bocquet, R., Equey Balzli, C., Gardet, E., & Gandia, R. (2021). SMEs' open innovation: Applying a barrier approach. California Management Review, 64(1), 113-137.
- Fleming, L., & Sorenson, O. (2001). Technology as a complex adaptive system: evidence from patent data. Research policy, 30(7), 1019-1039.
- Garzoni, A., De Turi, I., Secundo, G., & Del Vecchio, P. (2020). Fostering digital transformation of SMEs: a four levels approach. *Management Decision*.
- Gil-Gomez, H., Guerola-Navarro, V., Oltra-Badenes, R., & Lozano-Quilis, J. A. (2020). Customer relationship management: digital transformation and sustainable business model innovation. Economic research-Ekonomska istraživanja, 33(1), 2733-2750.
- Grossmann, I. E. (2012). Advances in mathematical programming models for enterprise-wide optimization. *Computers & Chemical Engineering*, *47*, 2-18.

- Iammarino, S., Piva , M., Vivarelli, M., & Von Tunzelmann, N. (2012). Technological Capabilities and Patterns of Innovative Cooperation of Firms in the UK Regions, Regional Studies, 46(10) 1283-1301
- Jansen, J. J., Van Den Bosch, F. A., & Volberda, H. W. (2005). Managing potential and realized absorptive capacity: how do organizational antecedents matter?. *Academy of management journal*, *48*(6), 999-1015.
- Jensen, R. J., & Szulanski, G. (2007). Template use and the effectiveness of knowledge transfer. Management science, 53(11), 1716-1730.
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255-260.
- Krüger, J. J. (2008). Productivity and structural change: a review of the literature. Journal of Economic Surveys, 22(2), 330-363.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.
- Leal-Rodríguez, A. L., Roldán, J. L., Ariza-Montes, J. A., & Leal-Millán, A. (2014). From *Potential Absorptive Capacity* to innovation outcomes in project teams: The conditional mediating role of the realized absorptive capacity in a relational learning context. International journal of project management, 32(6), 894-907.
- Moncada-Paternò-Castello, P. (2022). Top R&D Investors, Structural Change and the R&D Growth Performance of Young and Old Firms. Eurasian Business Review, 12(1), 1-33.
- Nelson, R. R. (2020). A perspective on the evolution of evolutionary economics. Industrial and Corporate Change, 29(5), 1101-1118.
- Organisation For Economic Co-Operation and Development (OECD) (2005). Oslo Manual. OECD, Paris.
- Nelson Richard, R., & Winter Sidney, G. (1982). An evolutionary theory of economic change. *Harvard Business School Press, Cambridge*.

- Pedron, C.D., Picoto, W.N., Colaco, M. and Araújo, C.C., (2018). CRM System: The role of dynamic capabilities in creating innovation capability. *BBR. Brazilian Business Review*, *15*, pp.494-511.
- Pellegrino, G. - Piva, M., (2020). Innovation, Industry and Firm Age: Are There New Knowledge Production Functions? Eurasian Business Review, 10(1), 65-95.
- Pittaway, D. S. (2004). Investment and second language acquisition. Critical inquiry in language studies: An international journal, 1(4), 203-218.
- Rabe-Hesketh, S. (2007). Everitt, Brian. A Handbook of Statistical Analyses Using Stata.
- Roberts, L. D. (2015). Ethical issues in conducting qualitative research in online communities. *Qualitative Research in Psychology*, *12*(3), 314-325.
- Saura, J. R., Ribeiro-Soriano, D., & Palacios-Marqués, D. (2021). From user-generated data to data-driven innovation: A research agenda to understand user privacy in digital markets. *International Journal of Information Management*, *60*, 102331.
- Savino, T., Messeni Petruzzelli, A. and Albino, V., (2017). Search and recombination process to innovate: a review of the empirical evidence and a research agenda. *International Journal of Management Reviews*, *19*(1), pp.54-75.
- Schoen, H., Gayo-Avello, D., Metaxas, P. T., Mustafaraj, E., Strohmaier, M., & Gloor, P. (2013). The power of prediction with social media. *Internet Research*.
- Schumpeter, J., (1911). The Theory of Economic Development. Cambridge, MA: Harvard University Press.
- Stata, A. (2015). Stata base reference manual release 14.
- Szulanski, G. (1996). Exploring internal stickiness: Impediments to the transfer of best practice within the firm. *Strategic management journal*, *17*(S2), 27-43.
- Vasile, M., (2021). IoT as a Central Disruptive Technology in the Development of Hyperconnected Business and Social Models. *Risk in Contemporary Economy*, pp.261-275.

- Von Tunzelmann N. (2009) Competencies and capabilities: a reassessment, Economia Politica 3, 435–464.
- Winter, S.G., (2003). Understanding dynamic capabilities. *Strategic management journal*, *24*(10), pp.991-995
- Yu, S. H. (2013). Social capital, absorptive capability, and firm innovation. *Technological Forecasting and Social Change*, 80(7), 1261-1270.
- Zahra, S. A., & George, G. (2002). The net-enabled business innovation cycle and the evolution of dynamic capabilities. *Information systems research*, *13*(2), 147-150.
- Zhao, Y., Von Delft, S., Morgan-Thomas, A. and Buck, T., (2020). The evolution of platform business models: Exploring competitive battles in the world of platforms. *Long Range Planning*, *53*(4), p.101892
- Zoia, M.G. - Barbieri, L. - Cortelezzi, F. - Marseguerra, G. (2018). The Determinants of Italian Firms' Technological Competencies and Capabilities. Eurasian Business Review, 8(4), 453-476
- Zuboff, S. (2015). Big other: surveillance capitalism and the prospects of an information civilization. *Journal of information technology*, *30*(1), 75-89.

Appendix 1: Jumpstart questionnaire

My age:

- 18-24 years
- 25-30 years
- 31-40 years
- 41-50 years
- 51-60 years
- >60 years

2. My highest education

- Primary education - Basic education in reading, writing and mathematics along with an elementary understanding of other subjects.
- Lower secondary education - Completion of primary level education, but not reaching completion in GCSEs.
- Upper secondary education - Completion of GCSEs or equivalents such as international baccalaureate.
- Post secondary education - Completion of A-level or equivalent level of studies or vocational studies.
- Tertiary education - These programmes may be academically based or practically oriented / occupationally specific. Includes Bachelor, Master, PhD.

3. In my function I carry managerial responsibility (executive, end responsible, partner)

- Yes
- No

4. My enterprise is active in the following branche/industry

- Agriculture, forestry and fishing
- Extraction of minerals
- Water extraction and distribution; waste and wastewater management and remediation
- Construction industry
- Wholesale and retail
- Transport and storage
- Rental of movable property and other business services
- Public administration, public services and compulsory social insurance
- Education
- Health and welfare services
- Culture, sports and recreation

○ Other services

○ Industry, specify:

5. In which region is your enterprise located?

[▾]

6. What is the postcode of your enterprise? [ ]

7. In which geographic markets does your enterprise sell goods and/ or services? *

☐ Local

☐ Regional

☐ National

☐ EU-wide

☐ Beyond EU

8. Please indicate your company`s age in years [ ]

9. How many FTE work in your organization? [ ]

10. What was your firm’s turnover in the last calendar year? (in €) [ ]

8. Please indicate your company`s age in years [ ]

9. How many FTE work in your organization? [ ]

10. What was your firm's turnover in the last calendar year? (in €) ☐

11. How much time was spent working on core processes? ☐

12. Innovative capacity

12.1 To what extent do you feel up-to-date with state of the art in your field?

☐ Very low extent

☐ Some extent

☐ Moderate extent

☐ High extent

☐ Very high extent

12.2 To what extent do you feel able to adopt data-informed processes into your work processes

☐ Very low extent

☐ Some extent

☐ Moderate extent

☐ High extent

☐ Very high extent

13. What is the share of turnover from products or services introduced in the last 2 years that were new to your company? (in %) ☐

14. Have you introduced or improved any of the below in the last 2 years?

	Yes	No
14.1 Methods of manufacturing for producing goods or services?	◌	◌
14.2 Logistics, delivery or distribution methods for your inputs, goods or services?	◌	◌
14.3 Supporting activities for your processes, such as maintenance systems or operations for purchasing, accounting or computing	◌	◌
14.4 New business practices for organizing procedures (i.e first time use of supply-chain-management, business re-engineering, knowledge management, lean production, quality management, etc.) using data-informed decision making?	◌	◌

15. To what extent have you heard about the following tools and techniques *

	Never heard	Heard of it once	I know something about it	We want to use it	We already work with it
15.1 Spreadsheets	◌	◌	◌	◌	◌
15.2 Structured Query Language (SQL)	◌	◌	◌	◌	◌

	Never heard	Heard of it once	I know something about it	We want to use it	We already work with it
15.3 Data visualization tools (Tableau, Kibana, PowerBI)	◌	◌	◌	◌	◌
15.4 Application Programming Interface	◌	◌	◌	◌	◌
15.5 Reinforcement Learning	◌	◌	◌	◌	◌

16. My organization analyses the following data: *

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
16.1 Product data (e.g., sales)	◌	◌	◌	◌	◌
16.2 Financial data (e.g., billing)	◌	◌	◌	◌	◌
16.3 Customer data (e.g., orders, contracts)	◌	◌	◌	◌	◌
16.4 Employee data (e.g., absence. Productivity)	◌	◌	◌	◌	◌

17. My organization has the following digital infrastructure: *

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
17.1 A central storage for all data (a data warehouse)	○	○	○	○	○
17.2 Spreadsheet software such as Microsoft Excel, Libreoffice, OpenOffice	○	○	○	○	○
17.3 Access to a cloud computing platform (AWS, Google, Azure)	○	○	○	○	○
17.4 Access to external data via APIs or scraping	○	○	○	○	○
17.5 Databases such as SQL Server or Oracle	○	○	○	○	○
17.6 Statistical software such as SPSS or Stata	○	○	○	○	○
17.7 Data visualization software such as Tableau, PowerBI or MapInfo	○	○	○	○	○
17.8 Programming languages such as R or Python	○	○	○	○	○

18. Within my organization we spend a lot of time on *

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
18.1 Gathering and digitizing data	◯	◯	◯	◯	◯
18.2 Managing and maintaining a database / datafiles	◯	◯	◯	◯	◯
18.3 Running data analyses (e.g., comparative statistical analyses, predictive analyses)	◯	◯	◯	◯	◯
18.4 Visualizing data (e.g., making graphs)	◯	◯	◯	◯	◯

19. Please let us know the extent to which you agree to the following statements when it comes to data within your organization: *

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
19.1 My colleagues often bring new ideas and developments with regards to data to the table					
19.2 My colleagues in general know their way around with new data related technologies					
19.3 My organization strives for fast adoption of innovations in the field of data					
19.4 When it comes to data, my organization has the means and opportunities to implement new developments quickly					
19.5 I am confident that the data within my organization is accurate					
19.6 I am confident that the data within my organization is up to date					
19.7 Support when running into issue with regards to data is well taken care off					

20. Please let us know the extent to which you agree to the following statements when it comes to data within your organization: *

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
20.1 My organization is aware of the possibilities of working with data	○	○	○	○	○
20.2 My organization likes to work with external parties when it comes to data gathering and analyses	○	○	○	○	○
20.3 My organization stimulates experimenting with new technologies	○	○	○	○	○
20.4 My organization often takes part in events with data as one of the main topics (e.g. Meetups, conferences, seminars)	○	○	○	○	○
20.5 My colleagues often bring new developments with regards to data to the table	○	○	○	○	○
20.6 When new data becomes available, I use this to review my opinion	○	○	○	○	○

21. Please let us know the extent to which you agree to the following statements: *

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
21.1 My organization is aware of the importance of data security	○	○	○	○	○
21.2 Within my organization everybody uses the same software (Office, Salesforce, Dynamics etc.)	○	○	○	○	○
21.3 My organization is aware of the law and regulations when it comes to data (e.g., the GDPR)	○	○	○	○	○
21.4 I am always strictly law abiding with the laws and regulations regarding data	○	○	○	○	○

Appendix 2: Instrumented Probit Regression results-The impact of FBD Data projects

	Model 1	Model 2
Dependent Variable completion		
Covariates		
Potential absorptive capability	*1.619****	
Realized absorptive capability		*-1.63****
Q7 - Geographic - Local	*.302**	*-.126*
Q7 - Geographic - Regional	*-.174*	*-.071*
Q7 - Geographic - National	*-.379***	*.102*
Q7 - Geographic - EU-wide	*.215*	*.05*
Q7 - Geographic - Worldwide	*-.141*	*.056*
Q15.1 - Tools - Spreadsheets	*-.067*	*.127*
Q15.2 - Tools - SQL	*-.006*	*-.001*
Q15.3 - Tools - Visualization	*-.066*	*-.092*
Q15.4 - Tools - API	*-.068*	*.065*
Q15.5 - Tools - Reinforcement Learning	*.007*	*.224***
Q16.1 - Analyzes – Product data	*-.068*	*.065*
Q16.2 - Analyzes - Financial data	*.118*	*.129*
Q16.3 – Customer data	*-.009*	*-.119*
Q16.4 – Employee data	*-.293****	*.226****
Q18.1 - Time - Gathering data	*-.269****	*.177**
Q18.2 - Time - Managing data	*-.009*	*.097*
Q18.3 - Time - Analyzing data	*.028*	*.053*
Q18.4 - Time - Visualizing data	*-.063*	*.108*
Q21.1 - Security - Importance	*-.146*	*.24***
Q21.2 - Security - Software	*-.239****	*.144*
Q21.3 - Security - Law	*.263***	*-.235**
Q21.4 - Security - Obey	*-.1*	*.158*
Q14.1 - Improved - Producing goods	*.198*	*-.021*
Q14.2 - Improved - Logistics	*-.103*	*-.076*
Q14.3 - Improved - Supporting	*-.241*	*.089*
Q14.4 - Improved - Organizing	*.472****	*.061*
Q17.1 - Infrastructure - Data warehouse	*-.087*	*-.002*
Q17.2 - Infrastructure - Spreadsheets	*.032*	*-.04*
Q17.3 - Infrastructure - Cloud	*-.004*	*.022*
Q17.4 - Infrastructure - External	*-.043*	*.091*
Q17.5 - Infrastructure - Databases	*-.066*	*-.005*
Q17.6 - Infrastructure - Statistical	*-.011*	*.061*
Q17.7 - Infrastructure - Visualization	*-.011*	*-.02*
Q17.8 - Infrastructure - Programming	*-.222***	*.155*
Q9FTE	*.005****	*-.003***
Q8Companyage	*.001*	*-.004*
Sectors		
Agriculture, forestry and fishing	*(base)*	*(base)*
Construction industry	*.026*	*.955**
Culture, sports and recreation	*.346*	*.758*
Education	*-.218*	*.806*

Health and welfare services	*.128*	*1.193***
IT	*.449*	*1.503***
Other services	*.256*	*.63*
Real estate	*(empty)*	*(empty)*
Transport and storage	*.305*	*.937**
Wholesale and retail	*.625*	*1.06***
_cons	*3.499****	*-5.874****
Observations	*247*	*246*
Standard errors are in parentheses		
****** p<.01, ** p<.05, * p<.1***		